DOCUMENT RESUME

ED 038 424	24	TE 001 810

AUTHOR	Marascuilo, Leonard A.; Loban, Walter
TITLE	An Empirical Study of the Dominating Predictive Features of Spoken Language in a Representative Sample of School Pupils: A Multivariate Description and Analysis of Oral Language Development.
INSTITUTION	California Univ., Berkeley.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO	BR-7-I-106
PUB DATE	69
GRANT	OEG-9-8-070106-0031(010)
NOTE	199p.
EDRS PRICE	EDRS Price MF-$1.00 HC-$10.05
DESCRIPTORS	*Behavior Change, Child Language, Conditioned Response, *Language Development, Language Fluency, Language Learning Levels, Language Patterns, Language Skills, *Language Styles, *Language Usage, *Oral English, Standard Spoken Usage, Verbal Development

ABSTRACT

To determine whether language behavior represents an early conditioned verbal response or whether it changes with age and experience was the purpose of this study which attempted to define unique isolates of language on the basis of actual language produced by young children. Tape recorded data were collected for 12 years from 211 children in Oakland, California. Data collected during the first three grades were used to define eight "language style groups" (research groups) and statistics recorded during grades 10-12 were used to assess and predict language facility and growth. To create the research groups, three test or rating variables (e.g., intelligence test and verbal performance scores) and 15 language variables (e.g., "average length of communication unit") were utilized. The basic hypothesis--children will not change with age their relative positions to each other in language behavior--was supported with respect to speech conventionality but not supported with respect to problems of mazes (groups of words not resulting in meaningful communication). It was supported with respect to fluency, dependent clauses, and elaboration index for students who began as poor users of oral English. These results have several implications for curriculum development, especially in the teaching of reading. (LH)

ED038424

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

BR-7-I-106
PA 24

AN EMPIRICAL STUDY OF THE DOMINATING PREDICTIVE FEATURES OF SPOKEN LANGUAGE IN A REPRESENTATIVE SAMPLE OF SCHOOL PUPILS:

A Multivariate Description and Analysis of Oral Language Development

Grant Number: OEG-9-8-070106-0031(010)

Project Number: 7-I-106

Leonard A. Marascuilo

Walter Loban

University of California
Berkeley, California, 94720

1969

TE 001 810

Department of Health, Education, & Welfare
Office of Education
Washington, D.C., 20202

ACKNOWLEDGEMENTS

As would be expected, a study of the scope and magnitude presented in this report required the expertise of a number of qualified assistants and researchers. In this case, the investigators were most fortunate.

For the preparation of the basic language data, special credit should go to the regular members of Professor Loban's research staff. These include Miss Janis Kujawa and Mrs. Leslie Lepinsky who extracted the language data from student folders and prepared it for the final data analysis. Special appreciation must be expressed to Mr. and Mrs. Arthur Williams who studied and evaluated the computer outputs used in this investigation to cluster students of similar language behavior and who then prepared [redacted] statistics that eventually provided the verbal descriptions of the various groups of students.

While Professor Loban's staff supplied the expertise on language and its development, statistical and computer skills were supplied by Professor Marascuilo's staff. In the beginning Mr. John Reed served as computer programmer, statistical assistant, and expert on the use of the Tryon Cluster Analysis Programs used in the initial period of the project. His duties were eventually assumed by Mr. Darshan Sacdeva who then conducted all of the remaining computer and statistical work for this project with great efficiency and skill.

The importance of the contributions made by Dr. Neil Timm and Dr. Carol Treanor to this study cannot be overestimated. Without their assistance it is quite possible

that this study might not have been attempted.

Dr. Timm prepared the Dr. Jeremy Finn computer package for multivariate analysis for use at the Computer Center of the University of California at Berkeley. These programs were obtained through the generosity of Dr. Finn at the time Dr. Marascuilo taught at the State University of New York, at Buffalo. This program was modified by Dr. Timm and renamed CALIF. It was used for the canonical analysis, principal component investigation, multivariate analysis of variance, and linear discriminant function analysis employed in this study.

Dr. Treanor obtained from the IBM Corporation, through the cooperation of her husband, the Friedmann-Rubin Clumping Program which she then prepared for use at the University of California Medical Center at San Francisco. This program was used to create the eight language groups which serve as the basic core of this investigation. Because of her willingness to travel to San Francisco, the final grouping of students was obtained with ease.

Finally, special credit should be extended to Mr. Thomas Little, who prepared the rough drafts of the manuscript from less than ideal first manuscripts and then prepared the final typed report you are about to read.

Table of Contents

	Page
Introduction and Research Hypothesis	1
Description of the Basic Language Variables	5
Discussion of Variables and Reduction of Their Number	13
Principal Component Analysis on the Language Variables at Grades One, Two, and Three	16
Summary Findings Based on the Principal Component Analysis	23
Canonical Analysis of the Five Reduced Language Variables	24
Summary Findings Based on the Canonical Correlation Analysis	37
Cluster Analysis Based on the Principal Component Scores of the Eight Variables at Grades One, Two, and Three	39
The Language Profiles of the Eight Groups of the Study According to the Average T'Scores for the First, Second, and Third Grade Data	49
Summary Comments Concerning the Eight Language Groups of the Study	68
Univariate Analysis of Variance for Equality of Mean Value of the Eight Language Groups	71
Multivariate Analysis of Variance for the Mean Vector of the Eight Language Groups	78
Multivariate Analysis of the Average T Scores for the First, Second, and Third Grade Data	86
Scheffé Type Analysis of the Differences Between the Mean Vector of the Eight Language Groups at Grades One, Two, and Three	100
Summary Comments on the Similarities and Differences Between the Eight Language Groups	109
Language Profiles of the Average T Scores for the Tenth, Eleventh, and Twelfth Grades	111
Summary Comments on the Later Age Profiles of the Eight Language Groups	126

Page

Multivariate Analysis of the Average T Scores for the Tenth, Eleventh, and Twelfth Grade Data 128

Scheffé Type Analysis of the Difference Between the Mean Vectors of the Eight Language Groups at Grades Ten, Eleven, and Twelve 141

Summary Comments on the Similarities and Differences Between the Language Groups at Grades Ten, Eleven, and Twelve 147

Statistical Analysis of the Language Profiles Employing Linear Discriminant Functions 148

Summary Findings Based on the Linear Discriminant Function Analysis 165

Interpretations and Recommendations 167

- Diagnosis 172
- Method 175
- Content 177
- Evaluation 181

References 192

A MULTIVARIATE DESCRIPTION AND ANALYSIS OF ORAL LANGUAGE DEVELOPMENT

Introduction and Research Hypothesis

That human beings vary enormously in their command of the spoken word is a matter of common observation. Proficiency--and even more, power--with oral language is an important aid to adequate, successful living, no matter by what values one judges. "Give me the right word and the right accent, and I will move the world," wrote Joseph Conrad, paying tribute to the power of language to influence thought, feeling, and action, both within oneself and in others.

Just as adults vary in their command of the spoken word, so too do children, and to foster language growth, schools need to avoid an inflexible regimen for all pupils. The single textbook curriculum to which all pupils are exposed in platoon or block fashion has always been one of the serious defects of American as well as world-wide education. Project Talent, a study of 440,000 American high school students, identifies the lack of effective procedures for individualizing instruction as the most serious defect in American education.(1) To individualize instruction completely creates formidable problems, especially in skills requiring a communication situation and therefore unsuited to programmed instruction. However, if pupils with similar language needs and difficulties could be identified, the task of organizing instruction could be carried out with greater precision and effectiveness. Determining whether or not such groups of pupils exist and describing their language development are the major purposes of the present research.

Of course, it is obvious that pupils' needs in language growth are not identical. In crude fashion, teachers can easily locate those whose speech varies for sociological reasons--foreign parentage or non-standard dialects such as Negro, Pidgin, or Appalachian. But even among those who speak standard informal English there are different needs based upon psychological rather than sociological features. For instance, do some speakers hesitate, and in halting fashion, fall into word tangles whereas some others express themselves fluently and easily? Do some pupils, regardless of dialect or the use of English as a second tongue, speak in organized, coherent fashion whereas others are confused in their thinking and muddled in their syntax? Do some know the full repertoire of syntactical patterns--such useful ones as appositives and infinitive clauses--yet seldom use them? Such clusters of similar language behavior, if they can be determined, should have important implications for planning more efficient and effective language instruction.

Research in language arts includes many investigators who focus on differences in pupils' language as one basis for explaining why pupils differ in their communication of thoughts, feelings, needs, and interests. Usually, the researcher describes the differences existing among children of different race, sex, social class, age, and other demographic and social factors. Much less frequently have researchers tried to explain differences in terms of language isolates and individual language styles though analysis in this direction is now beginning to emerge in the literature.

In this report a new approach will be described, one that attempts to define unique isolates of language on the basis of actual language produced by young children. Groups of children will be classified by their similarities in language behavior according to language variables observed at grades one, two, and three. On the basis of this classification and language data generated at grades 10, 11, and 12, an attempt will be made to show how these unique language behaviors either change and evolve with age to a new adult language style or remain fixed over time. As a research question, an attempt will be made to determine whether language behavior represents an early conditioned verbal response of each individual or whether it is fluid, changing with age and experience.

To develop analytical techniques, it will be hypothesized that language strategies are established early in life and undergo little change as a child increases in age. For instance, children who use complex elaboration in speech during their early years will be expected to continue this verbal behavior as they age, and on this language feature they will consistently maintain their same relative distance from subjects who use simple, unelaborated language and do not markedly improve their syntactical complexity as they age. Since language is not absolutely static but does change, usually improving in quality and complexity as age increases, it is really being hypothesized that children maintain their relative positions to one another as they age even though all subjects will generally show an improvement in language from childhood to adulthood.

The test of this hypothesis will be based on longitudinal data collected for twelve years on a selected representative sample of children living in Oakland, California, from 1953 to 1965. The present study utilizes 211 children whose spoken language was taperecorded at fixed yearly intervals under standardized conditions. The data from these tapes have been analyzed and the general characteristics of the variables have been discussed and reported by Loban in several earlier monographs.(2, 3, 4) The data collected from these 211 children during their first, second, and third grade tapings are used to define the language style groups that will become the research groups of this study. Statistics generated from the tenth, eleventh, and twelfth grade tapings are used to assess and predict language facility and growth.

The statistical procedures used for this analysis and testing are included in the broad framework of canonical correlation, principal components, statistical clumping procedures, multivariate analysis of variance, and linear discriminant analysis. However, rather than present a comprehensive review of these methods at this point, their basic properties will be discussed and their bearing upon the basic research question examined as they are introduced and needed.

Description of the Basic Language Variables

During the thirteen-year longitudinal study carried out by Loban, a large amount of information was collected on the language used by each of the subjects. From the original 328 subjects, complete data are available for 211 (e.g., for some of the 328 a test or a taping session is missing.) A thorough analysis of all of the accumulated data could take the collective lifetimes of a dozen researchers. Consequently, one of the goals of this present investigation is to examine some of the basic data and to reduce the massive amount of information to a relatively small set. This could reveal whether or not one can solidify and use as predictive measures the basic components of language and its development as measured by Loban's defined variables. Before undertaking such a study, one of the facts of research reality to be faced is that it might fail, but even if it does fail, much should be learned that would make further analysis simpler and more meaningful.

To create the observational research groups in this present study, it was decided to begin by using, (a) three test or rating variables, and, (b) a set of fifteen language variables generated during the first three years of the data collection period. These eighteen variables for all 211 subjects are the following:

Tests and Ratings

1. Mean intelligence scores as obtained from the Kuhlman-Anderson test of mental ability usually administered several times to each subject between grades two and seven.

2. Mean teachers' ratings of oral language performance during the thirteen-year period of schooling. (Each subject had 13 or more teachers over the entire study period.)

3. Verbal performance scores obtained from a kindergarten vocabulary test given at age five.

Language Variables Derived from Oral Language Taped Under Standardized Conditions

4. Fluency score one: average length of communication units at first grade. (A communication unit is each independent predication with all of its related modification.)

5. Fluency score two: average length of communication units at second grade.

6. Fluency score three: average length of communication units at third grade.

7. Fluency score four: freedom from mazes at girst grade.

8. Fluency score five: freedom from mazes at second grade.

9. Fluency score six: freedom from mazes at third grade.

10. Dependent clause usage: ratio of dependent clauses to communication units at first grade.

11. Dependent clause usage: ratio of dependent clauses to communication units at second grade.

12. Dependent clause usage: ratio of dependent clauses to communication units at third grade.

13. Conventionality index: success with use of standard English usage at first grade.

14. Conventionality index: success with use of standard English usage at second grade.

15. Conventionality index: success with use of standard English usage at third grade.

16. Elaboration index: amount of elaboration or complexity within the individual communication units at first grade.

17. Elaboration index: amount of elaboration or complexity within the individual communication units at second grade.

18. Elaboration index: amount of elaboration or complexity within the individual communication units at third grade.

Examples of these variables are shown below, with each example containing an extreme case at each end of the spectrum as well as a more average case from the center.

Examples of All Variables

Test and Rating Variables

1. Kuhlman-Anderson I. Q. scores range from 65 to 138 with a mean of 101.2 for the group of 211 subjects.

2. Teachers' ratings range from 1.8 to 4.3 with a mean of 3.2 for the group of 211 subjects. Note that each subject has one teacher's rating per year and the overal group mean is the average of cumulative means obtained on each subject over a thirteen-year period. The ratings were made by thirteen or more teachers.

3. Kindergarten Vocabulary Test scores range from 0 to 83 with a mean of 49.0 for the group of 211 subjects.

Language Variables

Before proceeding to the development of analytic techniques it will be useful to define carefully the terms used for the language variables in this research.

Communication Unit: A communication unit may be defined semantically as a group of words which cannot be further segmented (divided) without the loss of their essential meaning. Grammatically, a communication unit is any independent predication and all of its relevant modification. Thus, "I saw a man wearing a red hat" is a single unit of communication; if "wearing a red hat" were omitted, the essential meaning of that unit would have been changed and grammatically the participial modifier of man would be missing. Furthermore, "with a red hat" does not constitute a complete predication and it cannot stand alone. However, "I saw a girl and she was wearing a green hat" results in two communication units: (1) "I saw a girl"; (2) "[and] she was wearing a green hat." Dividing the sentence into two communication units does not result in loss of meaning to either unit and grammatically each is an independent unit. The average length of these communication units increases with advancing age, beginning with the brief sentences of very young children and progressing to the complex elaborated sentences of adults. The mean for the group of 211 subjects is 6.0 words per communication unit in grade one and 15.9 words per communication unit in grade twelve. Examples of communication units used by subjects in the oral language transcripts are as follows:

Short Units

Lower Grades: She is outside.
(3 words)

Upper Grades: He is plowing.
(3 words)

Medium Units

Lower Grades: They don't have very many clothes.
(7 words)

Upper Grades: And it is just about a father and his four boys.
(11 words)

Long Units

Lower Grades: Or we might play some games that I have in my house, some games that are in a box like that.
(21 words)

Upper Grades: And they're all working together to try to get her husband into this high political office to set him up for bigger and better things and maybe to become president or whatever he's got his mind on.
(40 words)

Maze Words as a Percentage of Total Words: A maze may be defined as a group of words or initial parts of words not resulting in a meaningful communication unit, i.e., a confused tangle of language not necessary to the communication unit. Most communication units contain no maze words whatsoever; thus the following examples are designed to illustrate the extent to which maze words can occur in a given communication unit.

Minor Maze Problem

Lower Grades: [and] and it looks like a cute little dog.
(1 maze word in a total of 9 words)

Upper Grades:	I think maybe the one that is running, the girl that is running, [knows] apparently knows something that the other one doesn't know because she's got sort of a puzzled look on her face. (1 maze word in a total of 36 words)

Moderate Maze Problem

Lower Grades:	[probably] probably [going] they're going back to their house. (2 maze words in a total of 10 words)
Upper Grades:	So what does Trina do but tell him to give her the money [which his last payoff] his last payroll because the company before firing him [had] of course [given him] had paid him off that money which he had deserved. (7 maze words in a total of 41 words)

Major Maze Problem

Lower Grades:	I got [one of my favorite toys a toy] my favorite toy in the garage. (7 maze words in a total of 15 words)
Upper Grades:	[and then and and] and [it's it's very] it's written effectively [so that] so that you think that [Leon-] Leonard's going to come in [and] and sort of you know [r-] release [his his] his love for Tolson [and his] and his need for Tolson [in] in this kind of weird relationship. (19 maze words in a total of 56 words)

Dependent Clause Ratio: A communication unit consists of an independent clause which may or may not be modified by one or more dependent clauses. Thus, "I saw a man" is an independent clause (as well as a complete communication unit) which may stand alone. One could also elaborate this with a dependent clause and produce "I saw a man who was wearing a scarlet hat" or with two dependent clauses and produce "I saw a man who was wearing a scarlet hat which was made of feathers." Actual examples from the oral transcripts of the subjects are as follows:

No Use of Dependent Clauses

Lower Grades:	I know that.
Upper Grades:	That's all.

Medium Use of Dependent Clauses

Lower Grades:	I don't know what that is. (1 dependent clause)
Upper Grades:	And it ended up the way she thought it would, somehow. (2 dependent clauses)

Large Use of Dependent Clauses

Lower Grades:	I think they're going home after a long day's work because it looks like it's getting to be night because the stars are out. (4 dependent clauses)
Upper Grades:	Well it was an illustration of how a man can be brainwashed to the point where he knows nothing but what he is told and does what he's told to do by a special person who's been set aside as his controller or master, however you'd like to put it. (6 dependent clauses)

Conventional English Usage: Standard English is defined as the type of language usage typically spoken in the political, social, economic, educational, and religious life of this country. This set of language habits is standard not because it is any more correct or capable than other varieties of English but rather because it is the type of English most frequently used in the conduct of the most important affairs of this country. Standard English ranges from informal to formal styles with many usages that are disputed or in transition. In this research, we are concerned with obvious departures from standard English usage, such as the deviant forms following:

Lower Grades: She aint got no dress on or nothing.
And the boy have a shirt on.
She don't know nothing.
And he brung it over.
And her is trying it.

Upper Grades: And this man and the horse was plowing.
Once upon a time there was two girls.
And then when they move into it, Marlene found out that she didn't like it because it too far from school.
And her mother and them liked it too.

Weighted Index of Elaboration: The weighted index of elaboration assigns specific numerical weights to the component syntactic elements within a communication unit. Thus, a unit with simple adjectives and adverbs as modifiers will receive fewer points than a unit containing more elaborated phrases or clauses; clauses or phrases embedded within other elaborated structures will receive still additional weight in this index. The following examples range from short, non-elaborated communication units to units containing a variety of embedded structures.

No Elaboration

Lower Grades: We play house.
(0 points)

Upper Grades: She was nineteen.
(0 points)

Medium Elaboration

Lower Grades: On Thursdays there's Deputy Dave again.
(2 1/2 points)

Upper Grades: And she's running towards it to see what's happening.
(14 1/2 points)

Extensive Elaboration

Lower Grades: Well that looks like there's an Eskimo travelling in a sleigh with a whole bunch of dogs pulling it.
(22 points)

Upper Grades: Well this isn't a plot so much as a situation where we'll say that the girl that's running beneath the tree is the daughter of the woman who's holding clothes or something in her hands. (27 1/2 points)

Discussion of Variables and Reductions of Their Number

As might be expected, these 18 variables possess many elements in common. This overlap of information can be seen by an examination of the correlation matrix or array reported as Table 1. In this table, the complete set of correlation coefficients for the 18 variables for the total group of 211 subjects is presented. Since $r_{xy} = r_{yx}$, only the correlations above the main diagonal of the correlation matrix are shown. These statistics are based on near complete data so that less than two percent of the raw data were estimated. Whereas the original Loban study consisted of 328 subjects, only those 211 on whom complete data were available were used in the present analysis.

Finally, it should be noted that variables corresponding to 4 through 18 measured at grades ten, eleven, and twelve are used to evaluate the effectiveness of the ability to predict later life speech patterns. These corresponding variables are:

19. Fluency score one: average length of communication units at tenth grade.

20. Fluency score two: average length of communication units at eleventh grade.

21. Fluency score three: average length of communication units at Twelfth grade.

Table 1. Matrix of Correlation Coefficients for the 18 Variables Relating to the Data for Grades One, Two, and Three. (Variable Numbers are as Listed in the Description of the Basic Language Variables.)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1.00	.54	.49	.49	.41	.36	.09	.17	.23	.30	.33	.27	.55	.49	.57	.43	.41	.30
2		1.00	.69	.43	.35	.41	.12	.26	.26	.32	.27	.35	.45	.48	.54	.41	.33	.34
3			1.00	.23	.21	.18	.25	.35	.41	.17	.29	.22	.46	.53	.59	.22	.20	.21
4				1.00	.71	.67	-.29	-.11	-.12	.59	.43	.40	.28	.26	.33	.79	.52	.38
5					1.00	.73	-.22	-.20	-.12	.42	.51	.39	.28	.22	.28	.57	.72	.40
6						1.00	-.17	-.07	-.14	.44	.39	.64	.22	.23	.25	.54	.59	.67
7							1.00	.55	.48	-.19	-.08	.00	.23	.17	.24	-.25	-.11	.02
8								1.00	.65	-.03	-.01	.03	.24	.23	.30	-.06	-.04	.09
9									1.00	-.11	.05	.01	.20	.23	.38	-.10	-.02	.00
10										1.00	.51	.42	.27	.26	.27	.79	.43	.41
11											1.00	.41	.27	.24	.27	.49	.68	.32
12												1.00	.24	.28	.28	.40	.45	.77
13													1.00	.70	.66	.29	.31	.18
14														1.00	.78	.29	.30	.27
15															1.00	.33	.36	.26
16																1.00	.52	.42
17																	1.00	.48
18																		1.00

22. Fluency score four: freedom from mazes at tenth grade.

23. Fluency score five: freedom from mazes at eleventh grade.

24. Fluency score six: freedom from mazes at twelfth grade.

25. Dependent clause usage: ratio of dependent clauses to communication units at tenth grade.

26. Dependent clause usage: ratio of dependent clauses to communication units at eleventh grade.

27. Dependent clause usage: ratio of dependent clauses to communication units at twelfth grade.

28. Conventionality index: success with use of standard English usage at tenth grade.

29. Conventionality index: success with use of standard English usage at eleventh grade.

30. Conventionality index: success with use of standard English usage at twelfth grade.

31. Elaboration index: amount of elaboration or complexity within the individual communication units, at tenth grade.

32. Elaboration index: amount of elaboration or complexity within the individual communication units, at eleventh grade.

33. Elaboration index: amount of elaboration or complexity within the individual communication units, at twelfth grade.

Thus, as can be seen, this study is based upon 33 variables measured on 211 independent subjects so that the total number of bits of information is given by 33 × 211 = 6,963. From these bits of data, the hypothesis of language use stability will be tested and evaluated.

Principal Component Analysis on the Language Variables at Grades One, Two, and Three

The basic research hypothesis of this investigation is that students do not change relative to one another with respect to their use of language as they progress from kindergarten to high school graduation. One way to evaluate this hypothesis is to examine the correlations between the language data generated during the first, second, and third grades with the corresponding information collected during the tenth, eleventh, and twelfth grades. Such an approach might not be especially fruitful, for if it were attempted, it would be necessary to examine the intercorrelations that exist between five language variables generated at each of the six grades. If, in addition, the three aptitude and achievement measures were to be included in this set, then the total number of unique measurements to be examined in the study of the intercorrelations increases to $5 \times 6 + 3 = 33$. With this many variables, the total number of distinct correlations to be examined is given by $\frac{33 \times 32}{2} = 528$, of which 153 were reported in Table 1. As might be expected, very few researchers are able to study this number of correlations and comprehend all of the information contained about the underlying variables and their intercorrelations. Thus, simplification is necessary.

As can be seen by an examination of the correlation matrix of Table 1 and the statistics reported in Table 2 on mean scores and standard deviations, the Loban data at grades one, two, and three shows remarkable, and exceptionally high, consistency. Whereas the average length of communication unit increases from 6.0 words at first grade to 6.9 words at third grade, the standard deviations remain remarkably constant: 1.4 to 1.3 words. Furthermore, careful examination of the correlation coefficients of the length of communication unit with the other variables of the study also demonstrates that the correlations are quite stable. For example, the average correlation of the length of communication unit with freedom from mazes is .16 with the range in correlations extending from .07 to .29. Thus, excluding for the moment the aptitude statistics on tests and ratings, the Loban language proficiency data show considerable consistency. The only exception for the language variables is the elaboration index which shows a monotonic increase with grade, going from 75.4 at grade one to 89.4 at grade three. However, the standard deviations remain constant and the correlations with these variables remain relatively fixed and constant across grades.

Such consistencies in variances and groupings of correlation coefficients suggest that the basic data contain a large amount of redundancy and that a reduction in data is possible and meaningful. As a first thought, one might feel that data reduction could be achieved by using only first grade results and ignoring and discarding the second and third grade statistics. Such a possibility is not necessarily

Table 2. Measures of Central Tendency and Variability for the 18 Variables Relating to Language Usage for the 211 Subjects of the Study at Grades One, Two, and Three.

Variable	Average	Standard Deviation
1 Vocabulary	49.0	15.9
2 Teacher rating of oral language	3.2	.6
3 I. Q.	101.2	12.8
4 Language fluency at first grade	6.0	1.4
5 Language fluency at second grade	6.5	1.4
6 Language fluency at third grade	6.9	1.3
7 Freedom from mazes at first grade*	7.3	4.0
8 Freedom from mazes at second grade*	6.6	4.0
9 Freedom from mazes at third grade*	6.0	3.7
10 Dependent clauses at first grade*	17.3	11.6
11 Dependent clauses at second grade*	20.3	12.5
12 Dependent clauses at third grade*	22.7	15.0
13 Conventionality at first grade*	3.9	3.0
14 Conventionality at second grade*	3.6	3.1
15 Conventionality at third grade*	3.2	2.3
16 Elaboration index at first grade	75.4	27.6
17 Elaboration index at second grade	81.6	24.8
18 Elaboration index at third grade	89.4	27.4

*Original score is multiplied by 100 to place the variable on a scale having 1 as a lower limit.

without merit and could be attempted. However, intuition suggests that perhaps all the data could be used to identify the major underlying variables of the matrix of correlations. In addition, it would seem reasonable that a variable based on three observations, such as the simple average, has greater reliability than any one measure for any one year. That this is indeed true is easy to show.

For any one of the language variables measured over three years, a general form for a composite measure is given by:

$$X = a_1x_1 + a_2x_2 + a_3x_3$$

where a_1, a_2, and a_3 are arbitrary weighting constants and x_1, x_2, and x_3 are the values of the variables at grades one, two and three, respectively. A statistical measure of the efficiency of this variable as a composite measure is given by its variance, which in this case is defined by:

$$\begin{aligned}\mathrm{Var}(X) = {} & a_1^2\mathrm{Var}(x_1) + a_2^2\mathrm{Var}(x_2) + a_3^2\mathrm{Var}(x_3) \\ & + 2a_1a_2\mathrm{Cov}(x_1,x_2) + 2a_1a_3\mathrm{Cov}(x_1,x_3) + 2a_2a_3\mathrm{Cov}(x_2,x_3)\end{aligned}$$

Since the variances and correlations of the Loban data are almost equal by sets, it makes sense to substitute the average variances and correlations into this equation. When this substitution is performed, it is seen:

$$\mathrm{Var}(X) = \sigma^2(a_1^2 + a_2^2 + a_3^2) + 2\rho\sigma^2(a_1a_2 + a_1a_3 + a_2a_3)$$

where σ^2 = average variance and ρ = average correlation. If $a_1 = a_2 = a_3 = 1/3$, then X is the simple average of the three years testing. For this set of numerical constants:

$$\mathrm{Var}(X) = \mathrm{Var}(\overline{x}) = \frac{1}{3}\sigma^2 + \frac{2}{3}\rho\sigma^2 = \frac{\sigma^2}{3}(1 + 2\rho)$$

As long as $\rho < 1$, it follows that $\mathrm{Var}(\overline{x}) < \sigma^2$, the variance of the variables for any one year and is therefore a more precise measure. As an example, it appears that for the statistics on the elaboration index reported in Tables 1 and 2, σ is approximately equal to 25, and ρ is approximately equal to 1/2, so that:

$$\mathrm{Var}(\overline{x}) = \frac{25^2}{3}\left[1 + 2\left(\frac{1}{2}\right)\right] = \frac{2}{3}(25)^2 = \frac{2}{3}\sigma^2$$

As this last result indicates, the variance of the average elaboration index is about 2/3 that of any one measure for any one grade level and in this sense is a much more precise measure of performance.

As this example suggests, the average value of the scores for three years is a better measure than any one of the individual years. When the correlation coefficients are exactly or nearly equal, this is indeed the best procedure for combining data. However, when the correlations are not exactly equal, such as in the present data, one can obtain a measure that is even better than the average of the three individual scores. This more efficient value can be found by Principal Component Analysis,(5) a process which consists of finding the best weighted sum of the variables. The procedure employed to determine the best weighting coefficients is very similar to the process shown in the previous discussion except that the values a_1, a_2, and a_3 are not pulled out of the air as was seemingly the case in the previous example. Instead, one starts with the Var(X) and chooses the weights so that the variance is maximized, subject to the conditions that $a_1^2 + a_2^2 + a_3^2 = 1$. When the variances and correlation coefficients are equal, the minimizing weights are given simply by $a_1 = a_2 = a_3 = \frac{1}{\sqrt{3}} = .58$. When this occurs, a researcher can, with no loss in generality, replace the individual a value by 1/3 and thereby produce the simple average. When the correlations are not equal, these coefficients are not equal and are much more difficult to determine. Fortunately, if the process has been programmed for a high speed computer, the determination of the optimum weights is a simple matter. So as to achieve this data reduction, a Principal Component Analysis was performed on each set of three variables shown in Table 1. The results of the analysis are summarized in Table 3. Since the values of the weights shown in Table 3 are almost all equal to .58, it is clear that the average value of the variables themselves could have been used in accomplishing the reduction of data.

So as to obtain uniform results, the Principal Component Analysis on the Loban data has been performed using the correlation matrix. This means that all variables have been transformed from their regular measurement scale to one with a mean value of zero and a standard deviation of one, or to one in which the mean is 50 and the standard deviation is 10. Since the basic research hypothesis of the study focusses on relative changes and not absolute changes, such a transformation does not affect the results. Also, since most researchers are used to the treatment of standardized scores in the study of behavioral data, it was decided to transform all scores to a mean of 50 and a standard deviation of 10. Essentially, this means that the average student in the group of 211 is the

Table 3. Principal Component Weighting Factors for the Five Sets of Language Variables Measured on Grades One, Two, and Three.

Variables	Characteristic	a_1: Grade One	a_2: Grade Two	a_3: Grade Three	Percent of Variance Explained by the Component
4, 5, 6	Language Fluency	.57	.59	.58	80
7, 8, 9	Freedom from Mazes	.54	.60	.58	71
10, 11, 12	Dependent Clauses	.59	.59	.55	63
13, 14, 15	Conventionality	.56	.59	.58	81
16, 17, 18	Elaboration Index	.58	.60	.56	65

referrent to which all comparisons must be made. While this may limit comparisons between studies, it contributes to internal controls in this study and it does not hinder the analysis of the basic research hypothesis.

Finally, it should be noted that use could have been made of the Principal Component scores for the three years and these numbers could have been employed for further data analysis in place of the averages of the first three school years. This would certainly have produced different results, but the differences would have been minor. In any case, the final results would have been virtually identical.

It should also be noted that the average values have a high degree of reliability as composite measures since each of them explains more than 63 percent of the variability existing in the original variable measurements. For conventionality, the principal component variable accounts for 81 percent of the total variability. For the average conventionality measure:

$$\mathrm{Var}(\overline{x}) = \frac{\sigma^2}{3}(1 + 2\rho) = \frac{\sigma^2}{3}[1 + 2(.7)] = \frac{2.4}{3}\sigma^2 = .80\sigma^2$$

so even this measure accounts for 80 percent of the variance. For behavioral data, these are quite high in numerical value.

Summary Findings Based on the Principal Component Analysis

Examination of the sample averages, standard deviations, and correlation coefficients of the 18 variables related to language use at grades one, two, and three suggested that the data contained a sufficient degree of redundency and communality of information. Because of this, a principal component analysis was performed on each set of language characteristics. It was noted that the weighting coefficients for each set were almost all equal to $a_i = \frac{1}{\sqrt{3}} = .58$, indicating that a simple average could be used to represent an individual student's language usage. Furthermore, the gain in precision obtainable from the mean score over an individual grade was large enough to warrant the mean score as an adequate measure of language performance.

Since the principal component analysis was based on the correlation matrix and not on the covariances, the averaging had to be based on standardized scores. As an example, for elaboration index, this standardization is defined by:

$$X = \frac{1}{3}\left[\frac{X_1 - \bar{X}_1}{S_1} + \frac{X_2 - \bar{X}_2}{S_2} + \frac{X_3 - \bar{X}_3}{S_3}\right]$$

$$= \frac{1}{3}\left[\frac{X_1 - 75.4}{27.6} + \frac{X_2 - 81.6}{24.8} + \frac{X_3 - 89.4}{27.4}\right]$$

Finally, to produce statistical measures that have a mean of 50 and a standard deviation of 10, the resulting averages were transformed to:

$$T = 50 + 10\left(\frac{X - \bar{X}}{S}\right)$$

for each language variable. Thus, the fifteen language variables were reduced to five. They are:

1. Fluency
2. Freedom from mazes
3. Dependent Clauses
4. Conventionality
5. Elaboration Index

Canonical Analysis of the Five Reduced Language Variables

The data reduction based upon the principal component analysis at grades one, two, and three was also performed for the data at grades ten, eleven, and twelve. Thus, 30 bits of information measured on each pupil have been reduced to 10. These 10 bits of information can be used to make a preliminary investigation of the truthfulness of the basic research hypothesis that students do not change relative to one another with respect to their use of language as they progress from kindergarten to high school graduation.

Even though 30 bits of information have been reduced to 10 bits, the number of intercorrelations for the 10 variables is given by $10 \times \frac{9}{2} = 45$, still a relatively large number. Just as variables can be reduced, correlations can also be reduced.

One way to simplify the study of a large number of correlations is to reduce the data by means of multivariate canonical correlations and canonical variantes.(6) While the computations involved in the determination of canonical correlations and canonical variates are extremely complex, an understanding of canonical correlations and what they measure is easy to acquire, and with the use of high speed computers their determination is simple.

To help the understanding of these multivariate measures, consider classical multiple regression theory. For this model, there exists a single univariate dependent variable Y and a set of p independent variables X: $(X_1, X_2, \ldots, X_p)$ which relate to Y collectively. From the set of independent variables, a linear compound of the following form is constructed: $L_X = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$, where $\beta_1, \beta_2, \ldots, \beta_p$ are arbitrary constants unspecified in advance of data collection. Next, the β_p are determined so as to give the best representation of the dependent variable Y. This is usually accomplished through the method of least squares in which the β_p so selected have the additional property that they maximize the correlation coefficient between Y and L_x. The best linear compound that accomplishes this task is called the multiple regression equation and the correlation coefficient between the estimated Y values and the observed Y values is called the multiple correlation coefficient of Y with X: $(X_1, X_2, \ldots, X_p)$ considered collectively.

For canonical analysis, the model is taken one step further in that Y is no longer univariate but is allowed to increase in dimension so that the canonical correlation model starts with X: $(X_1, X_2, \ldots, X_p)$ and Y: $(Y_1, Y_2, \ldots, Y_q)$ with $q \leq p$ and the pair of linear compounds being given by:

$$L_X = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$

$$L_Y = \alpha_1 Y_1 + \alpha_2 Y_2 + \ldots + \alpha_q Y_q$$

Next, the β_p and α_q are selected so as to maximize the correlation between L_X and L_Y. The resulting compounds are called canonical variables and the correlation coefficient between the compounds is called the canonical correlation.

If Y consists of only one variable, then the canonical correlation is identical to the multiple correlation and the single canonical variate is identical to the multiple regression equation. If $q \geq 2$, it can be shown that q different sets of linear compounds $\left[L_X^{(1)}, L_Y^{(1)}\right]$, $\left[L_X^{(2)}, L_Y^{(2)}\right], \ldots, \left[L_X^{(q)}, L_Y^{(q)}\right]$ exist for the two sets of X and Y variables. These sets of variables have the added property that the correlation coefficient between the sets is zero. Operationally, this means that if the underlying variables are jointly normal in form, then the information summarized in each canonical variable is statistically independent of the information contained in the rest. In addition, if the p canonical correlations are ordered by size from R_1, R_2, to R_q, they can be tested for statistical significance by a test derived by Bartlett.(7)

The pairs of canonical variates that are retained following a test of statistical significance are then examined for meaning by a subjective evaluation of the magnitudes of the individual estimates of the α_q and β_p and the correlations of the individual parent variables with the newly manufactured canonical variates. On the completion of this analysis, it is customary to give names to the resulting canonical variates and use them as hypothetical constructs for the remainder of a scientific discussion. This practice will be followed in this narrative. Hopefully, it will provide valuable insight into the relationships existing between the variables and also simplify the basic analysis of the 33 variables used in this study.

As is recalled, the basic decision reached following the principal component analysis of the 18 language variables for the first, second, and third grade data was to reduce the 18 variables to eight variables by replacing each set of fluency measures, freedom maze scores, dependent clauses scores, conventionality measures, and elaboration indices by averages of the first three years data. As reported, an identical data reduction scheme was also performed on the tenth, eleventh, and twelfth grade data. Thus, while 30 bits of language usage information were obtained for each subject of the study, a data reduction to 10 bits per subject has been performed. These 10 bits of language information along with the three test and rating measures constitute the basic data of the study.

Since correlation coefficients are invariant and unaffected by changes in averages and standard deviation, all reduced scores were further translated to a mean of 50 and standard deviation of 10. This means that certain analyses to be performed on the data must be interpreted in the light of this perspective. Whereas absolute differences might be of interest, this study focusses on relative differences. It is for this reason that the basic research hypothesis has been stated in terms of relative changes. Thus, it is possible that all 211 students became more conventional in their speech over the range of years covered by the study. This would imply that language changes did occur and the hypothesis of no language change should be rejected. However, if a student starts at the first, second, and third grades at 1.30 standard deviations above the average student, then the research hypothesis states that at the tenth, eleventh, and twelfth grade this same student will still be 1.30 standard deviations above the average student, even though the average student is .8 standard deviations above his initial standing at grades one, two, and three. In this sense, no change corresponds to no change in relative distance and standings. This way of viewing the interpretations must not be forgotten as this narrative is studied since it limits the kinds of interpretations that are justified.

The intercorrelations for the 10 bits of language information is shown in Table 4. Since the correlation matrix is symmetrical about the main diagonal, only the upper portion of the matrix is presented. As can be seen, the correlation matrix has been partitioned into four sub-matrices of correlations. This is done to facilitate the reading of the table. The intercorrelations for grades one, two, and three appear in the upper five rows and first five columns. The intercorrelations for grades ten, eleven, and twelve appear in the bottom five rows and the last five columns. The intercorrelation of grades one, two, and three with grades ten, eleven, and twelve appear in the upper five rows and last five columns. If F_1 is used to represent fluency at grades one, two, and three, with corresponding notations for the remaining variables, at the early ages it is seen that the correlation between fluency and mazes is given by $r_{F_1M_1} = -.21$. At the later ages, this same correlation has been reduced so that $r_{F_2M_2} = .02$. The correlation between the fluency scores at the early and later grades is given by $r_{F_1F_2} = .37$, while the correlation between freedom from mazes for the two time periods is given by $r_{M_1M_2} = .43$.

If the sub-correlation matrix for the early years is examined, it is seen that the fluency, dependent clauses, and elaboration index are strongly correlated. The correlations of fluency with dependent clauses is given by $r_{F_1D_1} = .66$, between fluency and elaboration index is given by $r_{F_1E_1} = .80$, and

Table 4. The Intercorrelation for the Reduced Language Variables Measured at Grades One, Two, and Three and Grades Ten, Eleven, and Twelve.

		Grades One, Two, and Three					Grades Ten, Eleven, and Twelve				
		Flu.	Mazes	Dep. Cl.	Conv.	Elab. Ind.	Flu.	Mazes	Dep. Cl.	Conv.	Elab. Ind.
Grades	Fluency	1.00	-.21	.66	.32	.80	.37	.10	.30	.32	.21
One,	Mazes		1.00	-.06	.33	-.08	.06	.43	.18	.33	.17
Two,	Dep. Clauses			1.00	.37	.82	.34	.11	.30	.35	.25
and	Conventionality				1.00	.40	.33	.16	.33	.75	.29
Three	Elab. Index					1.00	.39	.13	.31	.35	.27
Grades	Fluency						1.00	.02	.74	.39	.78
Ten,	Mazes							1.00	.15	-.12	.12
Eleven,	Dep. Clauses								1.00	.37	.85
and	Conventionality									1.00	.32
Twelve	Elab. Index										1.00

and between dependent clauses and elaboration index is given by $r_{D_1E_1} = .82$. Apparently, these three variables are measuring common elements of language usage, and logically they should vary in the same direction. On the other hand, it appears that mazes and conventionality are measuring unique characteristics of language since the intercorrelations with these variables are quite low.

Essentially, the same sort of correlations are noted for the later ages. The correlations of fluency with dependent clauses is given by $r_{F_2D_2} = .74$, between fluency and elaboration index is given by $r_{F_2E_2} = .78$, and between dependent clauses and elaboration index is given by $r_{D_2E_2} = .85$. As noted for the early ages, freedom from mazes and conventionality show little relationship with the remaining variables or with each other.

Finally, if the early data is compared with the later data, it is seen that only one correlation is high and that is the correlation of conventionality at the early ages with conventionality at the later ages. The value of this correlation is given by $r_{C_1C_2} = .75$. The remaining correlations are quite low.

In Table 5 are shown the weights to be attached to the five language variables at each time period for those canonical variates that are statistically significant. As can be seen, the first set of canonical variates has a canonical correlation coefficient given by $R_{(1)(1)} = .79$. The hypothetical constructs that possess this correlation are defined by:

Table 5. Canonical Variates and Correlations Between the Five Language Variables at Grades One, Two, and Three With the Five Language Variables at Grades Ten, Eleven, and Twelve.

Canonical Pair	One		Two		Three	
Value of R	.79		.41		.34	
Canonical Weights	$L_X^{(1)}$	$L_Y^{(1)}$	$L_X^{(2)}$	$L_Y^{(2)}$	$L_X^{(3)}$	$L_Y^{(3)}$
Fluency	.27	.04	.36	-.07	-.69	-1.44
Mazes	.34	.24	1.06	.93	.17	-.01
Dependent Clauses	.12	.17	-.02	.17	.03	-.52
Conventionality	.73	.87	-.83	-.44	.42	.46
Elaboration Index	-.08	-.06	.15	.06	-.38	1.16
Correlations with Canonical Variates						
Fluency	.45	.46	-.02	-.05	-.87	-.73
Mazes	.52	.37	.70	.91	.48	.08
Dependent Clauses	.48	.50	-.02	.15	-.59	-.42
Conventionality	.94	.96	-.31	-.27	.11	.07
Elaboration Index	.49	.43	.01	.12	-.75	-.26

$$L_X^{(1)} = .27T_{F_1} + .34T_{M_1} + .12T_{D_1} + .73T_{C_1} - .08T_{E_1}$$

$$L_Y^{(1)} = .04T_{F_2} + .24T_{M_2} + .17T_{D_2} + .87T_{C_2} - .06T_{E_2}$$

Examination of the coefficients or weighting factors shows that these two hypothetical constructs are remarkably alike for the two different time periods. This suggests that $L_X^{(1)}$ and $L_Y^{(1)}$ are measuring the same language characteristic at the two time periods covered by the study.

For the early years it is seen that the conventionality has the greatest weight with $\beta_{C_1}^{(1)} = .73$. Also, conventionality shows the greatest weight on L_Y with $\alpha_{C_2}^{(1)} = .87$. As can be seen by examining the correlations reported in the lower portion of Table 5, the correlation between $L_X^{(1)}$ and conventionality at grades one, two, and three is given by $r = .94$, while for the $L_Y^{(1)}$ and conventionality at grades ten, eleven, and twelve, $r = .96$. On the other hand, it appears that the remaining variables contribute little to the two canonical variates. However, such is not the case. If one examines the correlations of the two canonical variables with the five language variables, it is seen that conventionality is not the only variable defining $L_X^{(1)}$ and $L_Y^{(1)}$. The correlations with fluency, mazes, dependent clauses, and elaboration index at the early years with $L_Y^{(1)}$ are given respectively by .45, .52, .48, and .49, while at grades ten, eleven, and twelve, these correlations are given by .46, .37, .50, and .43. Even though the simple correlations of the canonical variates with fluency, mazes, and dependent clauses, and elaboration index are of moderate strength, the exceptionally large correlations with conventionality indicate that the first set of canonical variates are primarily measures of conventionality. As was noted, the simple correlation between conventionality at time one and time two is given by $r_{C_1C_2} = .75$. Since $R_{(1)(1)} = .79$ is just slightly larger than $r_{C_1C_2} = 75$, this further supports the hypothesis that $L_X^{(1)}$ and $L_Y^{(1)}$ are primarily measures of conventionality. Furthermore, since the canonical correlation between the two sets is so large, it suggests that the numerical values of $L_Y^{(1)}$ at the second time period can be predicted with considerable reliability on the basis of the numerical value of $L_X^{(1)}$, and that with respect to this variable the hypothesis of no change in language use is tenable.

The canonical correlation for the second pair of statistically significant canonical variables is given by $R_{(2)(2)} = .41$. The associated canonical variates are defined by:

$$L_X^{(2)} = .36T_{F_1} + 1.06T_{M_1} - .02T_{D_1} - .83T_{C_1} + .15T_{E_1}$$

$$L_Y^{(2)} = -.07T_{F_2} + .93T_{M_2} + .17T_{D_2} - .44T_{C_2} + .06T_{E_2}$$

This set of variates does not seem to have the stability over time that the first set of canonical variates have. However, it appears that both variates are defined mainly by freedom from maze measures. This conclusion is also supported by the correlations of the individual language variables with the two hypothetical constructs. At the early time period, the correlation of mazes with $L_X^{(2)}$ is given by $r = .70$, while at the later time period, the correlation of mazes with $L_Y^{(2)}$ is given by $r = .91$. Thus, it appears that $L_X^{(2)}$ and $L_Y^{(2)}$ are primarily measures of the use of mazes in spoken language.

The correlation coefficient for the third pair of canonical variables is given by $R_{(3)(3)} = .34$ with the two canonical variates defined as:

$$L_X^{(3)} = -.69T_{F_1} + .17T_{M_1} + .03T_{D_1} + .42T_{C_1} - .38T_{E_1}$$

$$L_Y^{(3)} = -1.44T_{F_2} - .01T_{M_2} - .52T_{D_2} + .46T_{C_2} + 1.16T_{E_2}$$

Examination of the associated correlations indicates that this third pair of variates is defined by fluency, dependent clauses, and elaboration index. Students who use language with ease and fluency, rich in dependent clauses and all forms of elaboration are using language with power and force. At the other extreme, students who are not fluent, avoid the use of dependent clauses, and speak in simple subject-predicate form without coloration are using language without force or power. Thus, the third hypothetical construct is really a variable measuring language power.

Thus, the five language variables measured at each of the two time periods give rise to a three factor theory of language use. The three factors appear to be conventionality, freedom from mazes, and language power.

As a further check into the no change hypothesis of language use, individual canonical analyses were performed on each set of language variables measured at the two time periods. Thus, when conventionality at grades one, two, and three was canonically related to conventionality at grades ten, eleven, and twelve, the resulting correlation was given by R = .78. Since this "pure" measure of association is so close to $R_{(1)(1)} = .79$, it further supports the conclusion that $L_X^{(1)}$ and $L_Y^{(1)}$ are mainly measures of language conventionality alone. This means that in the elementary school it is possible to predict conventionality scores at grades ten, eleven, and twelve with considerable accuracy and precision from just the conventionality scores at grades one, two, and three. It also suggests that with respect to conventionality, the research hypothesis of no change in language style is upheld.

For freedom from mazes, the canonical correlation of early years maze use with later years maze use is given by R = .47, suggesting that students who have serious maze problems at an early age are quite likely to have solved the problem before reaching high school graduation. However, some prediction of later years problems with mazes is possible even though it will not be overly strong or successful. With respect to mazes, the research hypothesis of no change is equivocal. Prediction is possible, but its level of precision is low. Since R = .47 is quite close in numerical value to $R_{(2)(2)} = .41$, it follows that $L_X^{(2)}$ and $L_Y^{(2)}$ are primarily measures of freedom from mazes.

For fluency, dependent clauses, and elaboration index, the "pure" canonical correlations of early language use to late language use are given by .38, .32, and .33, suggesting that with respect to these characteristics students undergo considerable changes as they grow and mature. Thus, the ability to predict later life success with language power on the basis of early years experience is not too promising inasmuch as the canonical correlations are quite low. Since these correlations are quite close in numerical value to $R_{(3)(3)} = .34$, it follows that $L_X^{(3)}$ and $L_Y^{(3)}$ are hypothetical constructs defined by fluency, dependent clauses, and elaboration index.

Summary Findings Based on the Canonical Correlation Analysis

On the basis of the canonical correlation analysis, it appears that three relatively stable language factors are measured by the five Loban language variables. They appear to be conventionality of speech, freedom from mazes, and language power. The numerical values of the canonical correlations for the hypothetical constructs defining these three hypothetical language variables from the observed language variables at grades one, two, and three with grades ten, eleven, and twelve are given by $R_{(1)(1)} = .79$, $R_{(2)(2)} = .41$, and $R_{(3)(3)} = .34$, respectively. The stability of these variables over time is indicated in the near identity of the constructs' weighting coefficients in relation to the five original language variables and to the similarity of the canonical correlation coefficients to the "purer" measures of correlation based on the analysis of each variable measured at time one with the same variable measured at time two.

Thus, as the analysis proceeds, one should expect to find conventionality and freedom from mazes appearing as unique independent characteristics of spoken language. Thus, knowing a person is conventional in speech in no way indicates that the speaker is or is not a user of mazes. The same statement applies for non-conventional speakers in that their use of mazes is unrelated to the way they use language.

On the other hand, fluency, dependent clauses, and elaboration index are not independent of one another; they will invariably covary together. Thus, a person who is fluent uses a language rich in elaboration of all kinds as well as dependent clauses. However, at the other extreme, minimal fluency will most likely accompany minimal elaboration and avoidance of dependent clauses.

Finally, the hypothesis of no change in language style will be supported with respect to conventionality. Students who start life with conventional speech will continue to use this mode of expression. On the other hand, students who start at the lower end of this language characteristic will continue at the lower end of this important style of expression. The canonical correlation of $R_{C_1C_2} = .78$ indicates that $R^2_{C_1C_2} = (.78)^2 = 60.84$ percent of the total variance at the later ages is predicatable from the early age performance for this variable. For behavioral data, this is a highly reliable result.

With respect to freedom from mazes, $R^2_{M_1M_2} = (.47)^2 = 22.09$ percent of the later age variance is predictable from the early age data. While some prediction is possible, it is not overly strong. For language power $R^2_{F_1F_2} = (.38)^2 = 14.44$ percent, $R^2_{D_1D_2} = (.32)^2 = 10.24$ percent, and $R^2_{E_1E_2} = (.33)^2 = 10.89$ percent, indicating that the ability to predict later life speech with respect to power is not too strong. There is some reason to expect this poor correlation over time in that students who use language with power at an early age are unable to increase their power because of the limits placed on the number of discrete words that can enter into a complete unit of thought, or if not of thought, then of reasonable communication to others. However, students who start speech with short, clipped, terse sentences have great opportunities as they age and mature to adopt powerful speech and gain experience with language at school, at home, and during social interactions with others. Thus, on this dimension, the bottom can rise sufficiently to match the top.

Cluster Analysis Based on the Principal Component Scores of the Eight Variables at Grades One, Two, and Three

Having reduced the early year's data matrix by substituting eight new variables conveying almost the same information as the original eighteen variables, the next task was to create the groups having similar language problems. The procedure used for this part of the research is based upon computer programs developed by H. P. Friedmann and J. Rubin of the IBM Corporation(8). The method used for this form of cluster analysis or clumping procedure is based upon a reverse form of the Wilk's Criteria for multivariate analysis of variance.(9, 10) For the Wilk's method, G multivariate normal populations are sampled and p variables (Y_1, Y_2, ..., Y_p) are measured on each of n_g ($g = 1, 2, ..., G$) individuals. On each of the p variables for each of the G samples, average values are computed to provide a "profile" on each sample. The hypothesis tested by the Wilk's method is that the G "profiles" for each of the universes from which the samples come are identical or when stated as a null hypothesis, the mean values of the G universes are equal, variable by variable. The test statistic for this test is quite simple in form and is termed Wilk's Criteria.

The evaluation of this statistic is based on two numbers called determinants. One determinant is based on the total dispersion for the observed data and the other determinant is based on the within dispersion for the observed data. The Wilk's Criteria is the ratio of these two measures. Under certain assumptions, this ratio can be referred to the F distribution by means of an approximation generated by Box.(11) With this transformation large values of the transformed Wilk's Criteria correspond to a rejection of the null hypothesis of identical profiles. To compute the Wilk's Criteria, the values of n_1, n_2, ..., n_G and the exact value of G, the number of groups to be compared, must be specified in advance of data collection. For the Friedmann and Rubin procedure these parameters are unknown and must be determined from the data. In this sense, cluster or clumping analysis represents a reverse multivariate analysis of variance.

In practice, the problem is solved by specifying the value of G in advance and then on a random basis assigning the individual subjects of the entire sample to one of the G groups. In this case G was set equal to eight and on the first computer run, the 211 subjects were randomly assigned to eight groups; then the Wilk's Criteria was computed. After the first trial was completed, one subject was taken from one group and transferred to another group, and the Wilk's Criteria was again computed and compared to the first value. If the first transfer improved the value of the Wilk's Criteria, the subject was left in place, a second subject moved, and the entire process was repeated. Following this second trial, the process was then permitted to run to completion, i.e., until the Wilk's Criteria could not be improved. Obviously, the number of trials was exceptionally large and the computer time required for this grouping or clustering of individuals is exceptionally high, extending to more than 20 minutes over 40,000 trials.

While the ideal clustering procedure would have made use of the eight new hypothetical variables, the computer costs for such a partitioning would have been prohibitive. For this reason, the three aptitude variables and the five reduced language variables at grades one, two, and three were submitted to a second principal component analysis and were thereby reduced to two variables. The factor pattern for these variables is as shown in Table 6 for both the unrotated and rotated factor patterns. For this rotation, the Varimax Rotation procedure of Kaiser(12) was employed since empirical research has shown that this particular rotation generally leads to interpretable and intuitively acceptable results.

As can be seen by examining the figures of Table 6, the final rotated variables can be characterized as measuring language power and language confidence. Factor One is mainly defined by measures of language fluency, dependent clauses, and elaboration index. Students who rank high on these three variables tend to be freeflowing language users who use

Table 6. Factor Patterns of the Two Hypothetical Variables Used for the Friedmann-Rubin Clustering Procedure.

Variable	Unrotated Factor Pattern Factor One	 Factor Two	Rotated Factor Pattern Factor One	 Factor Two
1. Vocabulary	.76	-.16	.45	.63
2. Teacher's Ratings	.78	-.27	.39	.73
3. Kuhlman-Anderson IQ	.68	-.53	.13	.85
4. Language Fluency	.74	.52	.90	.12
5. Freedom from Mazes	.20	-.76	-.37	.69
6. Dependent Clauses	.74	.45	.84	.18
7. Conventionality	.73	-.35	.30	.75
8. Elaboration Index	.80	.48	.91	.19
Eigen Value	3.93	1.78		
Percent of Explained Variance	49.2	22.3	36.7	34.7

language effectively whereas students who rank low on these three variables are more halting and laconic in expressing their thoughts and needs. Factor Two is mainly defined by measures of vocabulary, IQ, freedom from mazes, and conventionality. Students who rank high on these variables use language with confidence in a conventional manner while students who rank low on these variables use a non-standard speech. Thus, the two hypothetical variables used for the clustering analysis appear to be language power and language confidence.

The decision to generate eight language groups was based on a pre-trial analysis using Tryon's Cluster Analysis(13) which was available to the researchers early in the planning stages of this study. The Tryon procedures are based on the use of three variables which are originally partitioned into three equal-sized categories (low, medium, high) which produce twenty-seven possible groups of subjects. Once the groups are defined, subjects are shuffled from group to group in much the same way as in the Friedmann-Rubin program except that at each step a test is made to determine whether neighboring group clusters can be combined to form one group. With the 211 subject data, the original 27 groups were reduced to nine. Since one group contained only three subjects who were non-speakers of English, it was felt that eight language groups should be created. Thus, when data were presented to the Friedmann-Rubin program, the construction of eight different language groups was requested.

With the Friedmann-Rubin program, three different analyses were made. For the first analysis, it was thought that because the Tryon program did as well as it could to cluster individuals, the Friedmann-Rubin program could take, as its initial start, the output from the Tryon program and then attempt an improvement of it. For the second analysis, the Friedmann-Rubin recommendation to start with a random assignment of subjects was made. The third analysis was based on a compromise between the Tryon and the random start. By careful examination of the grouped subjects, it was finally decided that the most meaningful grouping was obtained from the random start. The reasons for this decision are based upon the recommendations of Mr. and Mrs. Arthur Williams, the major liason and coordinating researchers involved with the Loban study and exceptionally knowledgeable about all 211 subjects.

In addition to the determination of the Wilk's Criteria, the Friedmann-Rubin program also provides a roster of the 211 students according to their group membership. This listing was examined by Mr. and Mrs. Williams for each of the computer runs. On the basis of their extensive knowledge of the students with respect to speech patterns, school achievement, race and social class inclusion, they made the final decision that the random start clustering gave the most meaningful and interpretable grouping of the students. Pertinent comments and a summary of their evaluations are reported in Tables 7 and 8. From their summary, an attempt is made to define the language styles of the students comprising each group. In Table 9, the mean standardized score for each of

Table 7. Demographic Data on the Eight Groups Selected by the Random Start.

		Sex		Race*			IQ		Teacher's Rating**		Socio-Economic Status ***						
Group	Number of Subjects	M	F	C	O	N	Median	Range	Median	Range	1	2	3	4	5	6	7
One	27	17	10	22	0	5	112	100-125	3.87	3.04-4.82	6	6	11	1	1	2	0
Two	51	25	26	44	2	5	114	96-135	3.75	2.67-4.67	14	14	6	6	8	3	0
Three	24	6	18	17	3	4	110	91-128	3.44	2.66-4.29	2	9	5	3	4	1	0
Four	37	13	24	17	1	19	99	81-111	3.12	1.89-4.11	3	4	4	5	8	9	4
Five	10	4	6	0	0	10	88	83-94	3.03	2.38-3.79	0	0	1	0	5	4	0
Six	30	16	14	4	3	23	92	79-104	2.98	2.00-3.88	0	0	2	5	7	10	6
Seven	16	8	8	2	0	14	80	68-103	2.30	1.58-4.00	0	0	0	1	3	8	4
Eight	16	10	6	2	8	6	93	57-113	2.03	1.54-2.78	0	1	0	0	8	4	3

*C = Caucasian, O = Oriental, N = Negro

**1 = Low, 5 = High.

***1 = High, 7 = Low.

Table 8. Comments [redacted] Concerning the Characteristics of Students Clustered Together Under the Random Start.

Group	Comments
One	The best of our Superior Language Ability group is here. Slightly male. Caucasian. High socio-economic status. The Negro students are excellent. Many of these students were accelerated. All college material. High fluency, clause ratio, and elaboration index; slightly below average maze problem; below average usage problem.
Two	The good members (plus a few other outstanding ones) of our Superior Language Ability group are here. Evenly divided sexually. Caucasian. Above average socio-economic status. Some were accelerated. All college material. Above average fluency and elaboration index; slightly above average clause ratio--possibly indicating heavier use of other types of elaborated usage to cause the index and clause ratio to differ somewhat; below average maze problem and usage problem.
Three	Above average group in the basis of non-language (oral) variables. Good IQ, TR. However, these were non-talkers as small children and even as high school students in most cases. Many are very shy; the Orientals had a poor command of English. Mainly Caucasian with some Oriental and Negro representation. Mainly female. Above average socio-economic status. Below average fluency, clause ratio, and elaboration index; below average maze problem and usage problem. All these factors on the profile are strongly influenced by the lack of talk.
Four	Average group. Good racial representation. Strongly female. Socio-economic status ranges to all ratings. A few members of our Low Language Ability group appear here as well as one High Language Ability. These students graduated from high school, but few continued (unless highly motivated.) Slightly above average fluency, clause ratio, and elaboration index; average maze problem and usage problem. None of these students had what I would call a "stimulating" background.
Five	Average group. All Negro. Even sexual representation. Below average socio-economic status. IQ and TR place this group lower than the oral language scores. The scores for grades one, two, and three differ considerably, with grade one consistently highest. The profile is rather confusing in that these are all big talkers--lots of volume. This element accounts for their above average oral language scores on fluency, clause ratio, and elaboration index; a big maze problem and some difficulty with conventional usage.

Table 8. (Continued)

Group	Comments
Six	*Below Average* group which contains several members of our Low Language Ability group. Evenly divided sexually. Mainly Negro. Below average socio-economic status. The boys in this group do not speak well at all--many are shy; the girls are much more verbal. Some of these students have been in trouble. Several did not finish high school. Several of the girls had babies which made it even more difficult for them to attend school. *Below average* fluency, clause ratio, and elaboration index; *slightly above average* maze problem; *above average* usage problem.
Seven	This group is very *poor*. Many are in our Low Language Ability group. Negro mainly. Evenly divided sexually. Low socio-economic status. Nearly everyone in this group has had big problems in school and with the authorities. There is only *one* possibility here for education beyond high school. *Below average* fluency, clause ratio, and elaboration index; *huge* maze problem; *above average* usage problem. In the early years most of these subjects speak very poorly--nearly incomprehensible in that the mazes and meaning of the final unit are often garbled.
Eight	This group is *very low* on oral language. Slightly male. Strongly Oriental (50% compared to 5% for the total group). Low socio-economic status. The one high socio-economic status subject is a boy who just can't talk. On IQ, though not on TR which places this group lowest in the set, the group is higher than the three groups I placed above it. On language variables, however, the members score very low. *Extremely low* fluency and elaboration index; *below average* clause ratio--these are lower than any other group by far; *below average* maze problem--probably because they don't talk; *extremely high* problem with usage--probably because no one could understand them. The Orientals could only speak a few words of English; the other subjects are our lowest in language ability.

Table 9. Table of Mean Standardized Scores on Language Profiles for the Eight Language Groups Created by the Friedmann-Rubin Clustering Program.

Variable	Fluency	Freedom from Mazes	Dependent Clauses	Conventionality	Elaboration Index
Group					
One	60.5	51.3	62.0	55.8	61.9
Two	53.5	53.2	51.6	56.3	52.3
Three	41.2	57.3	44.0	55.4	42.6
Four	52.4	48.0	51.5	50.3	52.1
Five	60.2	38.6	60.0	46.2	58.5
Six	45.2	48.3	43.6	42.8	44.9
Seven	49.1	36.3	46.2	39.7	47.4
Eight	32.4	56.0	39.6	37.7	35.8

the variables and groups are presented. These means are used to help the defining of the language groups.

The Language Profiles of the Eight Groups of the Study According to the Average T Scores for the First, Second, and Third Grade Data

In this section, the students who comprise each of the language groups are scrutinized for similarities and differences with respect to school achievement, attitudes, demographic characteristics, family and home life, and language use. As such, much of the discussion is subjective and can be attacked with some justification as being non-scientific or non-objective. However, the investigators are amazed as to the particular students who were clustered into specific groups. Personal knowledge of some of the pupils at grades one, two, and three indicates that certain students grouped together actually belonged together at those early ages. For this reason, it is believed that true language isolates have been created which reflect the five language characteristics investigated in this study.

Certainly, the addition of more language variables to the clustering procedure is bound to change the clustering in that more language groups would be generated. The eight groups of this section serve only as a starting point for this kind of research. Whether or not it should be continued and expanded will have to wait until reactions to this study have been received and evaluated. Originally, the study was conducted to design and provide analytical techniques for further research, but as will be seen, the results are so consistent that the conclusions and recommendations require comment and further study so as to assess their importance to education.

The average standardized scores for each variable for each group is as shown in Table 9. The corresponding profiles, based on these averages, are shown graphically in Figures 1 through 8.

Group One. Subjects with broad-based vocabularies, superior, fluent, conventional English, and effective complex syntactical structures.

Of the twenty-seven students comprising this language group, 22 are Caucasian and five are Negro. At an early age the Negroes in this group demonstrated an excellent proficiency with spoken language and continued to show this command of oral language at later ages. The same observations characterize the Caucasian students who help define this language group. Twenty-three of the students come from the highest SES groups of the study, and on a rating scale of 1 to 5, their median teacher's rating of oral language proficiency is 3.87. All have a potential for higher education; as shown by their high median IQ of 112 points, nearly one standard deviation above average, it can be already predicted that they should be successful in completing four years of advanced education.

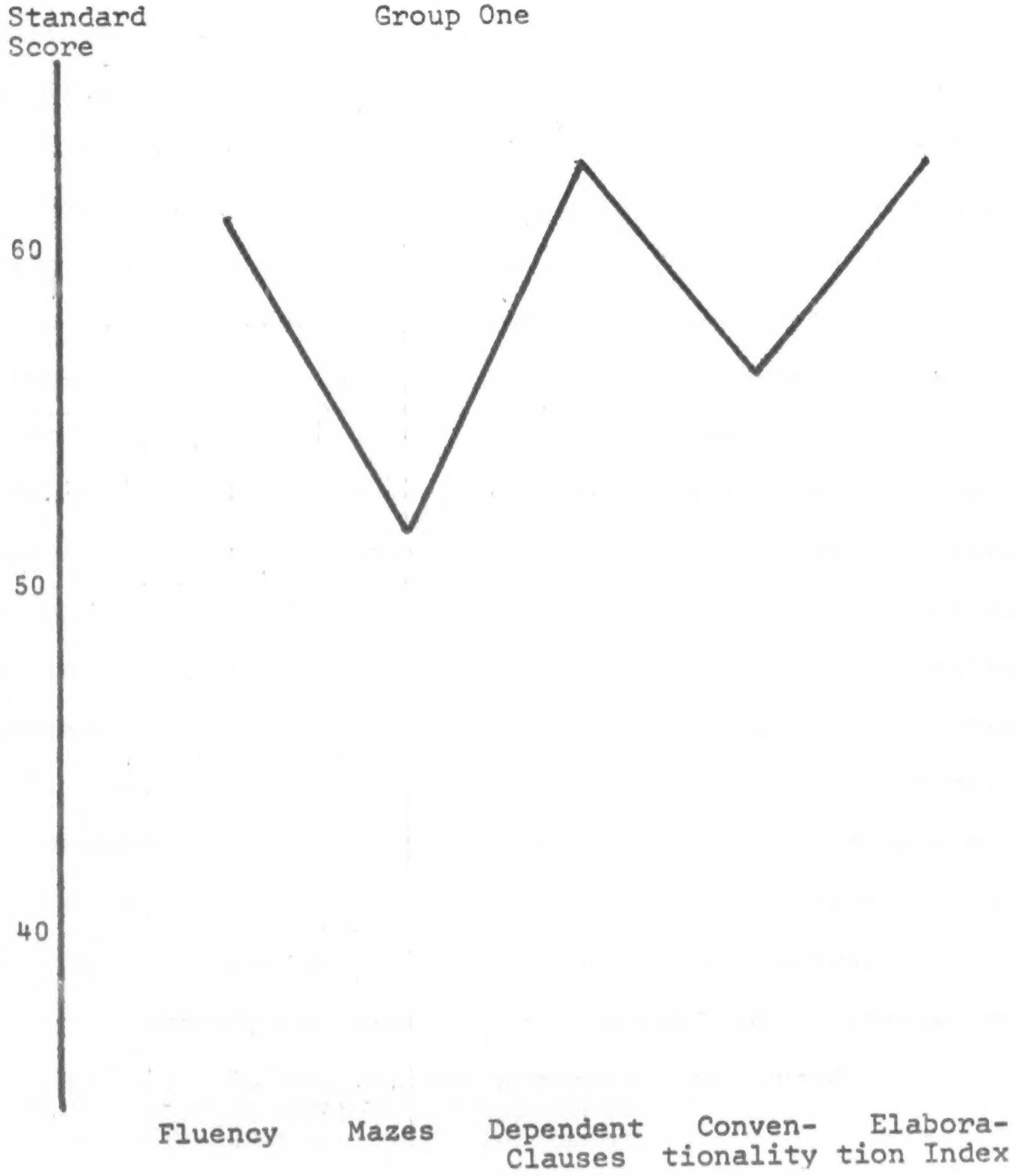
According to their first, second, and third grade language measures, their childhood use of oral language is impressive. Considering their age, these students are exceptionally vigorous in expressing thought; their speech is effective even though it is elaborate and complex; they have a remarkable repertoire of syntactical structures; they also tend to experience a slightly below average problem with mazes.

The profile for these students on the five basic language variables is shown in Figure 1. This profile is based on the statistics generated during their first three years of schooling. As is recalled, the scales of the five language variables are not in the same metric, and so it was decided to standardize all scores by conversion to a mean of 50 and a standard deviation of 10. On this standardized scale of measurement it is seen that the students of this group tend to average more than one standard deviation above the entire group on measures of fluency, use of dependent clauses, and speech elaboration. They tend to be average maze users and to be more conventional than most children in their usage. There is little doubt that these students are superior in language; they have both the ability and the skill to use effective complex syntactical structures and well-organized expression. Furthermore, they do this with more dexterity, clarity, and fluency than most children of their age. If one examines the language used in their home environment, it is seen that almost all these students are accustomed to hearing language of power and conventional acceptability used by parents, older brothers and sisters, and friends.

Group Two. Students who use English in a highly conventional, fluent, coherent, but not remarkably superior mode of expression.

This is the largest of the eight language groups generated by the clustering of students according to oral speech patterns used at the first, second, and third grade. This group of language users, equally divided between boys

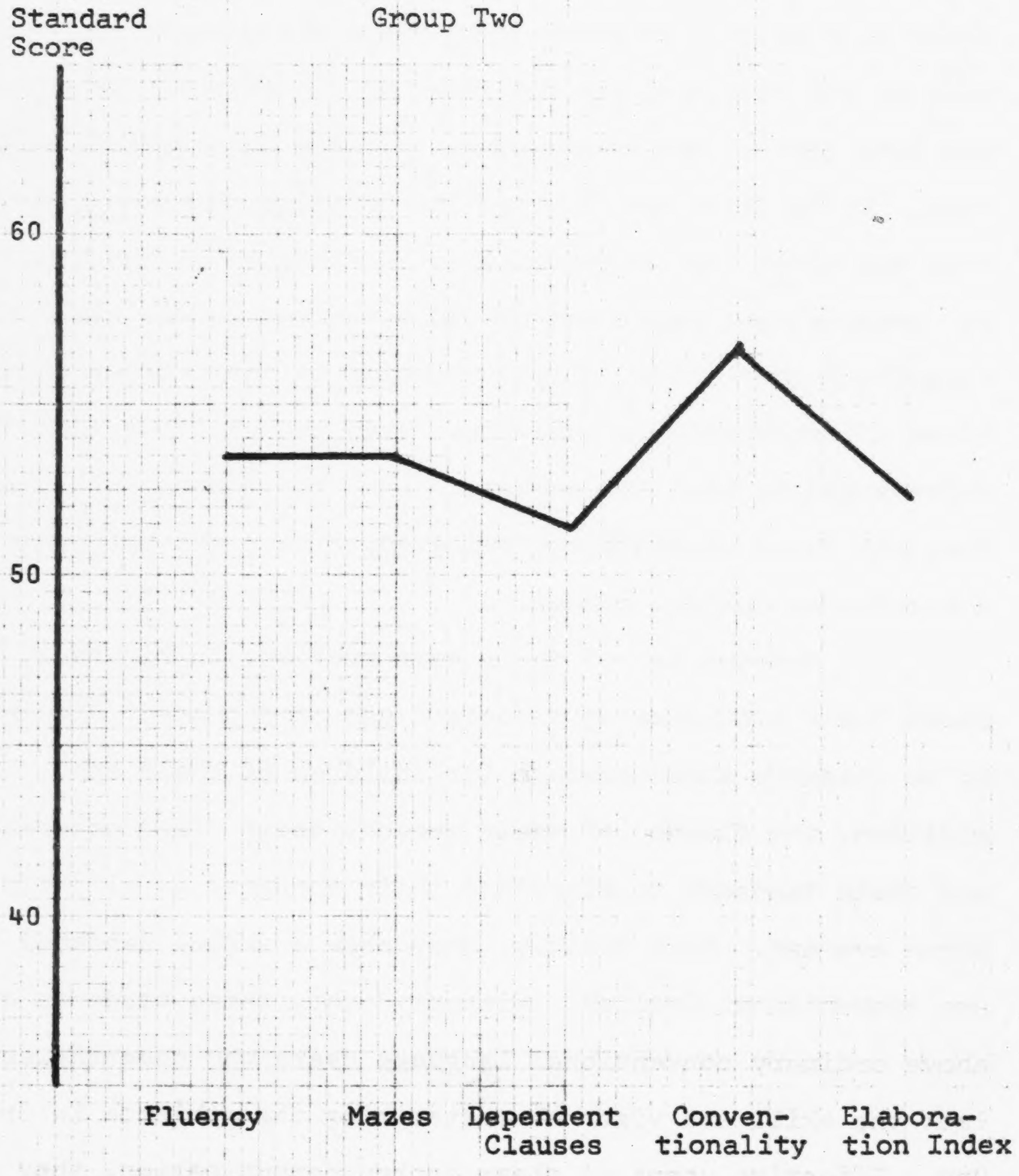
Figure 1. Subjects with Broad-based Vocabularies, Superior, Fluent, Conventional English, and Effective Complex Syntactical Structures.



and girls, is composed mainly of Caucasians. Their median IQ of 114 is quite high, and in line with their demonstrated intelligence, it is logical that teachers tend to give them a relative high oral language proficiency rating of 3.75. As might be expected, students with these characteristics adjust well to the middle-class structure of the American school of the late 1960's, and their school records support this supposition. While 34 of the 51 students defining this group come from the three top social classes of students in the study, 11, nonetheless, come from the lowest three social classes, suggesting that social class inclusion is not a major determiner of membership in this group even though it is markedly represented by high SES members. Like the members of Group One, all these students are believed to have the potential for a successful college career.

Examination of their mean profile, in Figure 2, shows their avoidance of mazes and use of dependent clauses to be slightly above average for children of their age. In addition, the fluency of their speech, their clear coherence and their tendency to elaborate their thoughts is slightly above average. Most notably, they show a strong tendency to use conventional English. Without doubt, these students are above ordinary conventional language users who nevertheless lack the skill and vigor characterizing the subjects in Group One. Effective users of clear spoken communication, they do not put into their speech the power characteristic of Group One. They experience no difficulties in fluency or conventionality. They are impressive in ability to communicate, but they are

Figure 2. Students Who Use English in a Highly Conventional Fluent, Coherent, but not Remarkably Superior Mode of Expression.

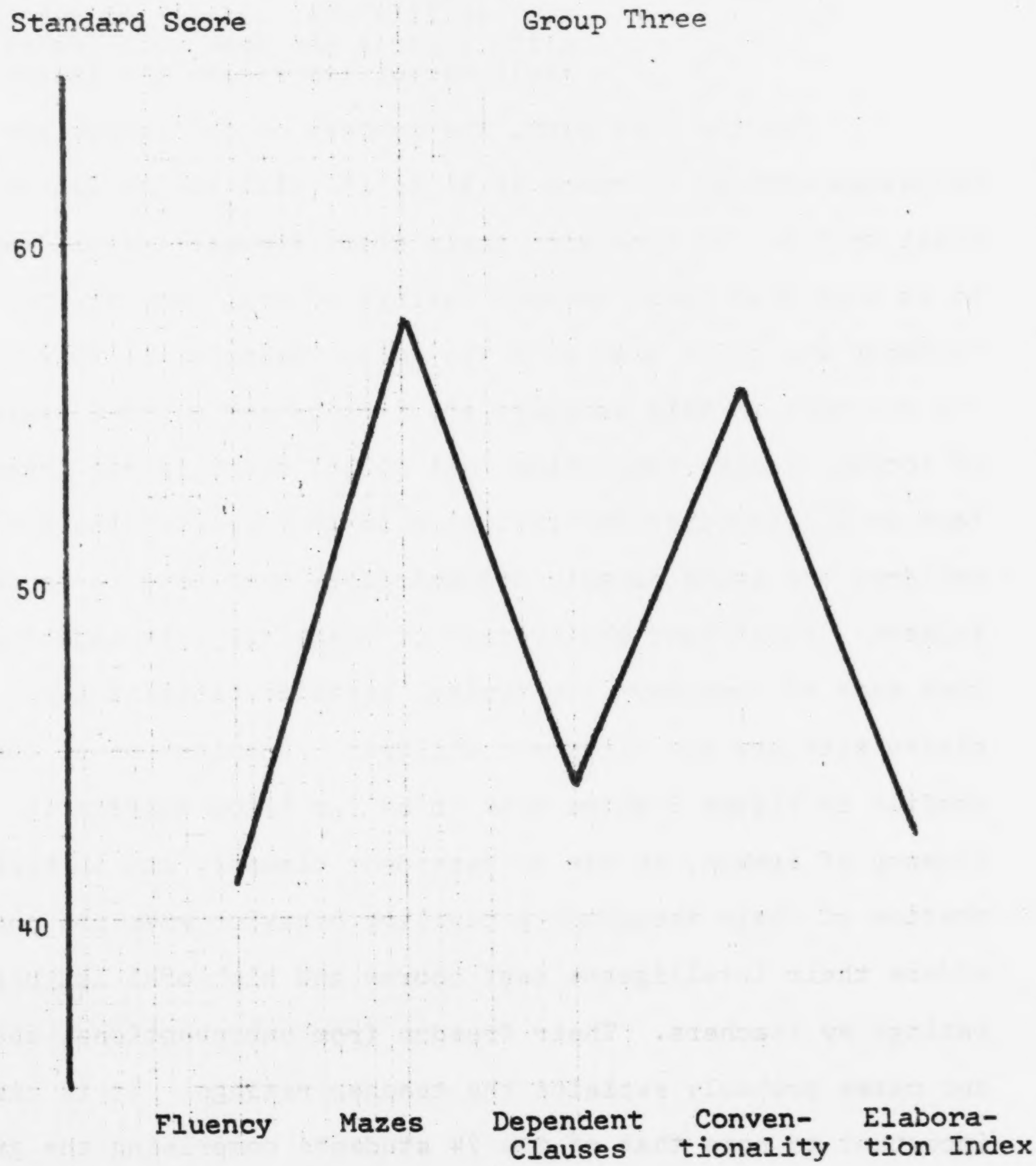


not outstanding. Their use of language, successful and rewarding, lacks color and texture. They lack the remarkable effectiveness and power characteristic of Group One.

Group Three. Students of above average intellectual ability who, because of personality shyness and weak self-images, avoid verbal expression and interchange.

For the most part, the members of this group are Caucasian with an IQ range of 91 to 128 with the median being equal to 110. In line with their above average intelligence, it is seen that their teacher ratings of oral language proficiency are quite high with the median being equal to 3.44. The subjects in this language group represent a broad range of social classes suggesting that social class is not important as a determiner for inclusion in this group. While these children are quite capable academically, they tend to be non-talkers. First-hand observation of these subjects indicates that many of them have the typical behavior patterns associated with shy and withdrawn children. Examination of their profile in Figure 3 shows them to be far below average in fluency of speech, in use of dependent clauses, and in elaboration of their thoughts--a puzzling behavior when one considers their intelligence test scores and high oral language ratings by teachers. Their freedom from unconventional usage and mazes probably explains the teacher ratings. It is also important to note that of the 24 students comprising the group, 18 are girls. Basically, these children come from home backgrounds where parents demonstrate considerable concern for the educational success of their children.

Figure 3. Students of Above Average Intellectual Ability Who, Because of Personality Shyness and Weak Self-images, Avoid Verbal Expression and Interchanges.



Both logical and empirical evidence suggest that personality and self-image have a major influence on language use and growth. In the case of these students, it appears that these factors tend to restrict both the child's experience with language and success with it. It is possible that these children could have shown language profiles similar to those of Group One or Two provided that they had parents similar to those of Group One or Two and provided that they had used language in greater quantity or volume. It would appear that a child who uses language effectively and is encouraged in this direction to use it often is quite likely to experience an exponential blossoming of vocabulary and oral language facility that tends to feed upon itself and continues to grow. The child who does not use language has a low probability of experiencing this geometric growth in language performance. The children in this group are obvious members of this latter category; they did not use language easily and frequently and so their exclusion from Group One or Two is understandable. For the most part, one can say that these are students of above average ability who avoided the use of language because of personal shyness, psychological insecurity, Oriental culture, or some similar inhibiting factor.

Group Four. Typical middle-class users of American English.

Of the 37 subjects comprising this language group, 17 are Caucasian and 19 are Negro. Their median IQ is 99 and their median teacher rating is 3.12. The members of this group cover all socio-economic classes and all of them could be typified as students of average ability. This is also true of their oral language as shown in Figure 4, where their average profile on the five language variables is demonstrated. On all of these variables these students tend to be very average with a mean standardized score close to 50. They are competent, conventional, average talkers. They comprise an eminently average set of subjects, an excellent measure of what to expect in oral language of typical American public school pupils. In only one respect do they exhibit any feature that is unusual--in this case, three-fourths of them are girls.

Group Five. Talkative subjects from culturally deprived backgrounds who use non-conventional speech patterns.

All ten members of this language group are Negro. Their IQ scores vary within a very narrow range of 83 to 94. It should be noted that their apparent low IQ scores are not necessarily indicative of lack of ability but may mainly reflect their basic unfamiliarity with written conventional English. These students come from low SES homes and as might be expected, they definitely show some difficulties in using conventional English. As indicated by their average profile, shown in Figure 5, they tend to average one-half a standard deviation below the average member of the entire sample of 211 students with respect to the use of conventional English. Yet they tend to be exceptionally fluent in their mode of speech; they use many dependent clauses; even though they have considerable problems with mazes, they express their ideas with lengthy elaborations. They have a lot to say when they are encouraged to speak out. The similarity of their profile

Figure 4. Typical Middle-class Users of American English.

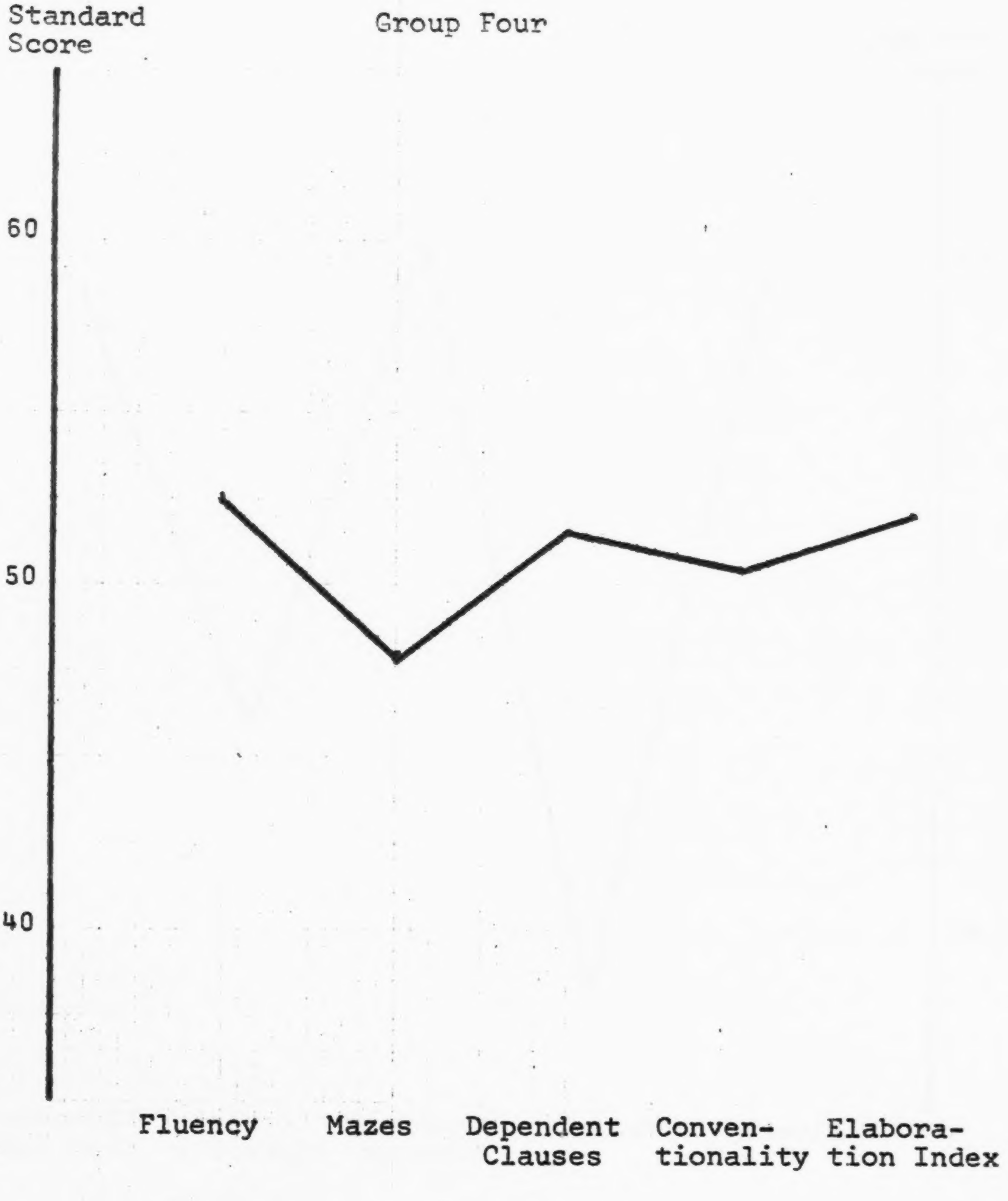
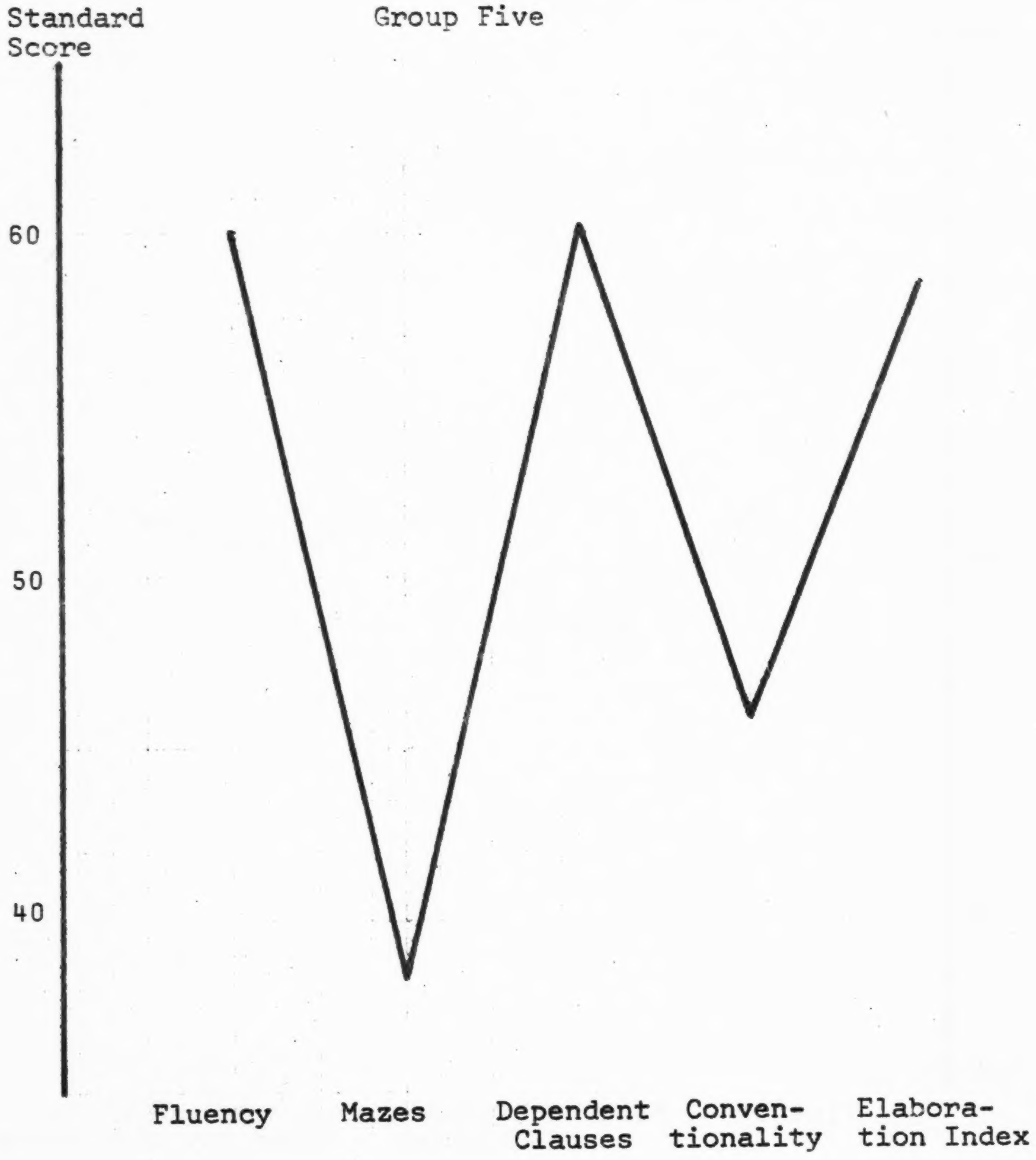


Figure 5. Talkative Subjects from Culturally Deprived Backgrounds Who Use Non-conventional Speech Patterns.



of students in Group One is striking and thought-provoking.

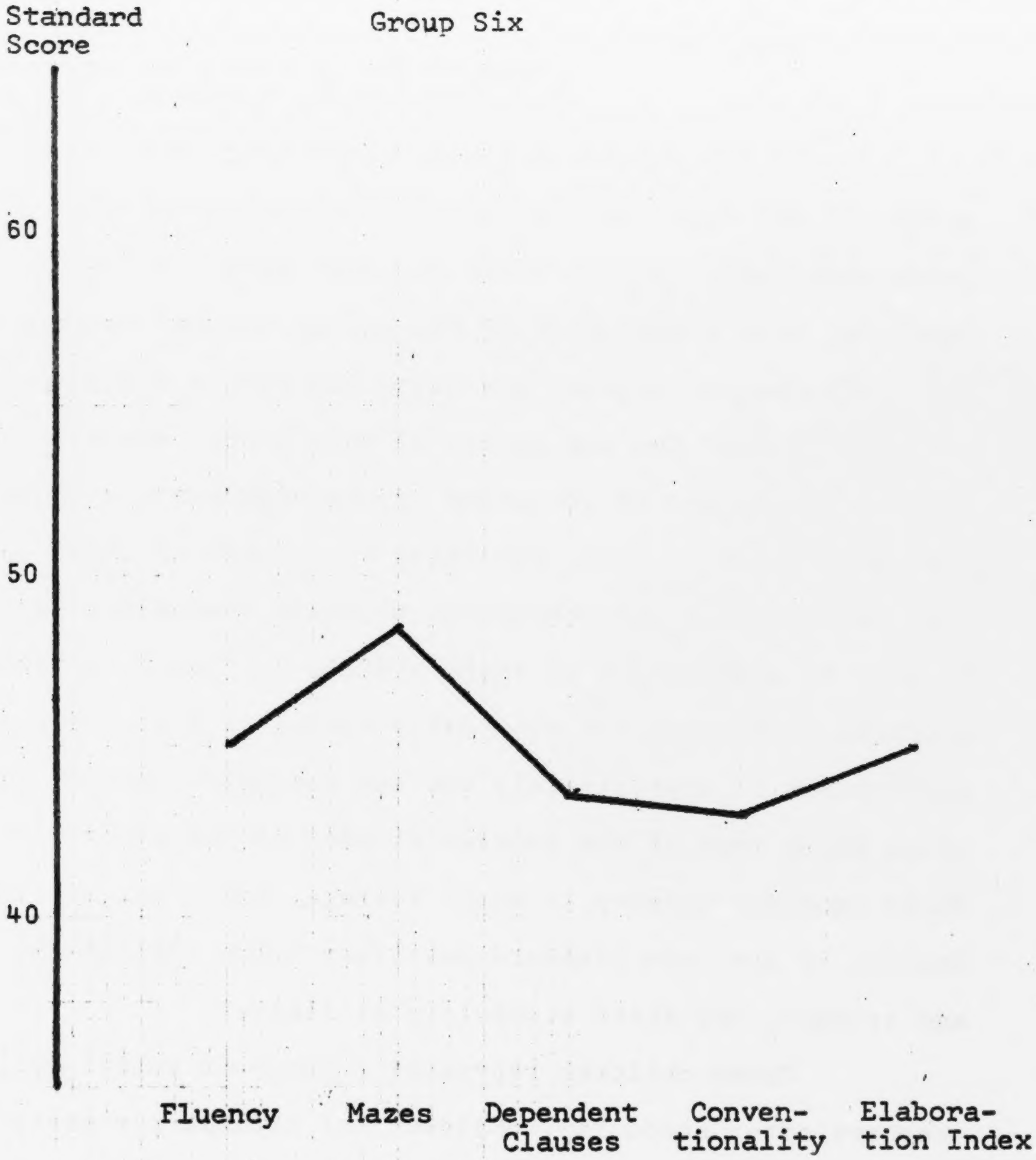
For many reasons this group is of unusual interest. Because of their superiority in fluency, as well as in their ability to phrase thought with complexity, they resemble the most superior group in oral language. Yet coming from minority group families, they lack the necessary exposure to appropriate standard usage needed for its acquisition. It may well be that this lack causes, especially in such able children, an uncertainty of thought and confusion of expression resulting in a greater number of mazes than is desirable for effective communication. One cannot help but speculate that if these students were given the advantages of a superior education in a non-discriminating culture, they might prove to be scholars and leaders of the community. While this speculation tends to ignore their below average IQ scores, it should be recalled that group intelligence tests are primarily verbal in nature and highly related to language ability. These children, being unfamiliar with standard English, would not be expected to do well on a paper and pencil IQ test. This suggests that these ten Negro children may very well represent pure examples of what is meant by disadvantaged pupils. Because their backgrounds are so different from middle-class school students, their language should be different. It is conceivable that the future of students similar to these could be significantly changed if their basic powers of oral language were identified and recognized early in their school careers. For the most part, these ten can be characterized as able, talkative speakers with non-conventional dialects. Their early exposure to spoken language occurred in an environment usually referred to as culturally deprived and disadvantaged.

Group Six. Children to whom standard English is a second dialect or a foreign language.

Of the thirty subjects who make up this language group, 23 are Negro and three are Oriental. Several of the remaining Caucasians manifest the peaked appearance and lassitude of manner often associated with "poor whites." Their median IQ is 92 and their median teacher's rating on oral language proficiency, 2.98, is that of average subjects. For the most part, these children come from low SES families. Twenty-three are from the lowest three socio-economic categories.

Examination of Figure 6 shows them to be about one half a standard deviation below the average for the entire sample in fluency, use of dependent clauses, conventionality, and elaboration. Their conventionality score is quite low, being almost one standard deviation below the average, indicating that these students are accustomed to speaking some form of non-standard English. Their basic problem, attempting to use standard English as a second language in school, conflicts with the dialect that they normally hear at home. Even in their own dialect they have a restricted syntactical repertoire. American education has not yet become sufficiently sophisticated linguistically to help these children improve their oral communication skills in anything more than a hit or miss fashion. The teachers of these children are well-intentioned but they operate in a situation where even the

Figure 6. Children to Whom Standard English is a Second Dialect or a Foreign Language.



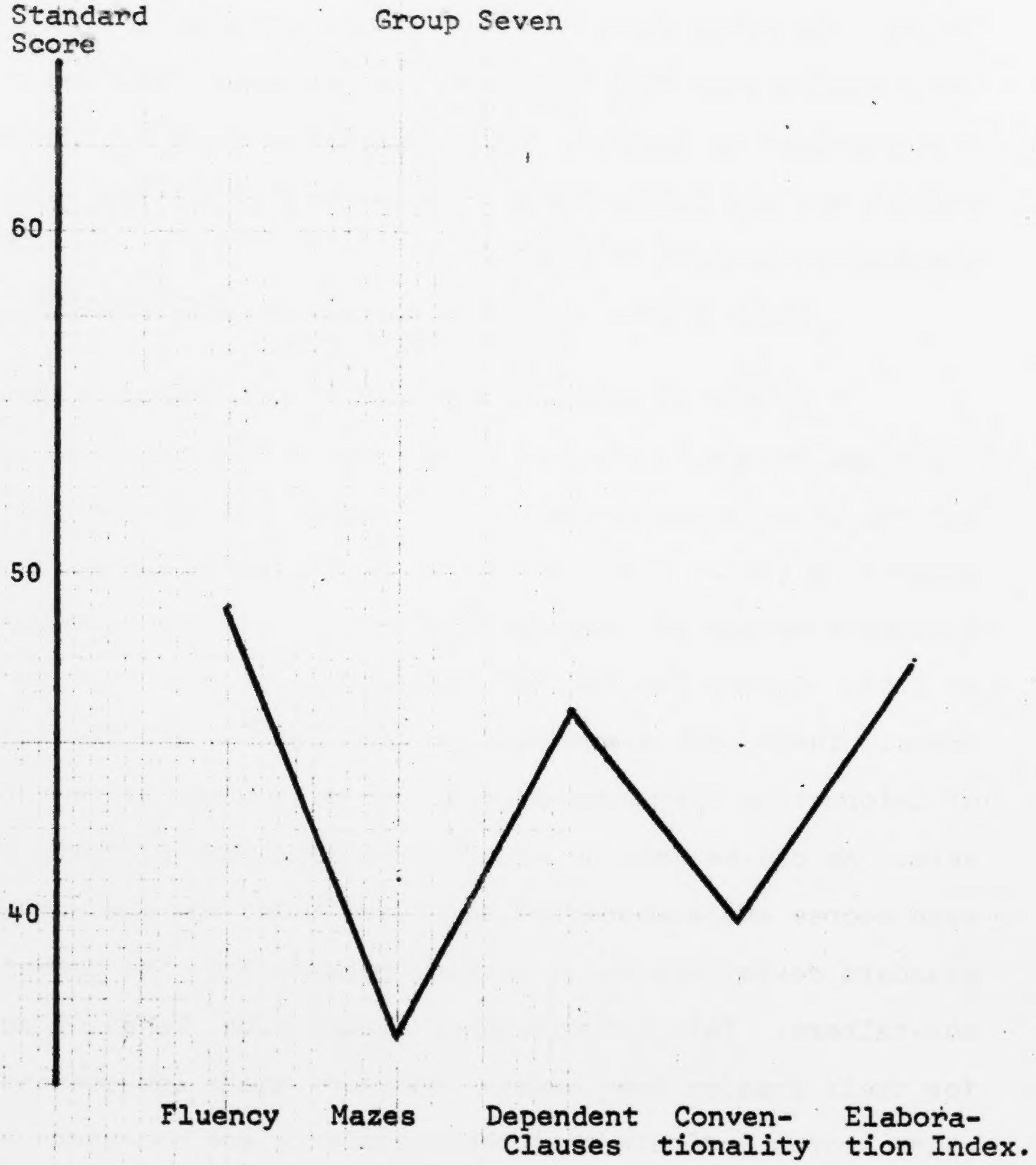
basic issues are unresolved, not to mention the lack of valid methods and programs. All these students can be summarized as language users for whom standard English is a second dialect or a foreign language.

Group Seven. Pupils whose use of English is non-standard and difficult to comprehend--for middle-class teachers.

Of the sixteen subjects comprising this language group, 14 are Negro and two are Caucasian (one of whom is Latin American). All of these children come from low SES families, have a median IQ of 80, and on teacher ratings of oral language proficiency are quite low with a median rating of 2.30. Except for one member of this group, there is probably no likelihood of education beyond high school. During the early years of their schooling the speech of these children was nearly incomprehensible to their teachers. As can be seen by examination of their profile in Figure 7, their problems with mazes are abnormally acute, in that their mean performance is approximately one and one-half standard deviations below that of the average student of the study. While their language fluency is about average, their use of standard English is low, one standard deviation below that of the average student, and their vocabulary is limited.

These children represent a group of school children teachers often encounter, children not notable for mental energy or ambition who find themselves ill at east in typical middle-class American schools. The two Caucasian pupils included in this group, as well as many of the Negro children, come from homes in which there was an ambivalent feeling

Figure 7. Pupils Whose Use of English is Non-Standard and Difficult to Comprehend--for Middle-class Teachers.

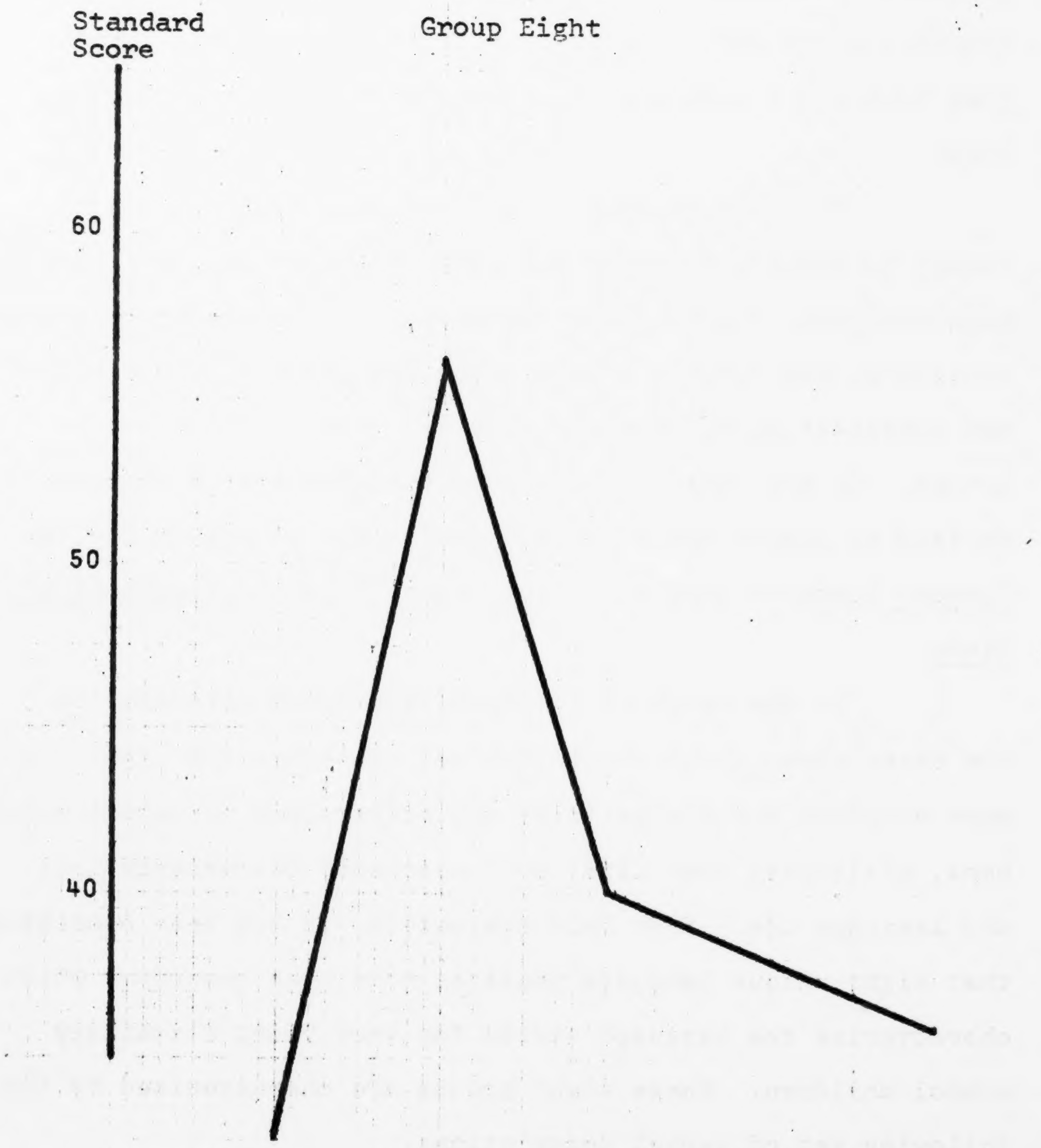


towards the pleasures and joys of having children. Some of these children were probably unwanted by their parents. Perhaps, as a result, these children manifested qualities of apathy and listlessness, as well as somewhat damaged self-images. In turn, these characteristics could well affect their proficiency with language. In any case, they can be characterized as limited, often incoherent users of standard English and not particularly able users of their own non-standard variations of English.

Group Eight. Students who are unconventional non-language users.

Of the 16 subjects who define this language group, eight are Oriental, six are Negro, one is Mexican-American, and one is a Caucasian whose own parents acknowledge his limited IQ (57). Their median IQ is 93, and their median teacher's rating of language proficiency is exceptionally low at 2.03. Except for the one Caucasian, all come from low SES homes. Their oral expression is very limited and the amount of information they express with labored speech is next to zero. As can be seen by examination of Figure 8, their fluency scores and elaboration index are about one and one-half standard deviations below average because they are essentially non-talkers. This inarticulate, almost mute, behavior accounts for their freedom from mazes. However, their low conventionality scores indicate that these students are non-users of standard English when they do decide to speak. Their inadequacy with language is striking. Their high freedom from mazes is solely the result of saying so little and saying that

Figure 8. Students Who are Unconventional Non-language Users.



in short units of talk. Eight of them come from Oriental homes where fluency is not a commendable trait and where Chinese, rather than English, is spoken. The others, both Caucasian and Negro, appear to be limited human beings both intellectually and in terms of vitality. In simple terms, they just don't have any go. Nothing moves them or excites them.

Finally, it should be noted that there is ample reason to believe that two distinct groups of children have been combined in this group--because of common language characteristics, not because of the known demographic, sociological, and intellectual differences existing between the two subgroups. In any case, all of these children can be characterized as inarticulate non-standard users of spoken English.

Summary Comments Concerning the Eight Language Groups of the Study

On the basis of extensive knowledge possessed by the researchers about the individual subjects, the eight groups were examined for similarities and differences on school achievement, attitudes, home life, socio-economic characteristics, and language use. From this evaluation, it has been concluded that eight unique language isolates have been generated which characterize the language styles for very young elementary school children. These eight groups are characterized by the following set of verbal descriptions.

Group One: Conventional English Speakers Who Use the Spoken Word with Power

These students are exceptionally vigorous in expressing thought. Their speech is highly effective, elaborate, and complex. They have a remarkable repertoire of syntactical structures, tend to be average users of mazes and use standard English more accurately than other children.

Group Two: Conventional English Speakers Who Use the Spoken Word in an Unremarkable Way

This, the largest of the eight language groups, includes children who tend to avoid mazes and dependent clauses in their speech. Even so, their fluency and tendency to elaborate on thoughts is above average for children of early school ages. The most notable characteristic about them is their tendency to use conventional English for expression. Their use of language is successful and rewarding, but on the other hand, lacks the color, texture, and imagery of Group One.

Group Three: Conventional English Speakers Who Avoid Verbal Expression

The children comprising this language group tend to be non-talkers. They are shy and withdrawn and come from homes in which parents expressed considerable concern for the educational success of their children. For the most part, these children are girls who perform well in school but do not use verbal interchanges to assist their learning and success in school. One wonders whether or not their high ratings from elementary school teachers may be the result of good behavior rather than language power.

Group Four: Typical Middle-class Users of American English

These children are competent, conventional, average users of American English. They serve as an excellent measure of what to expect in oral language of typical American public school pupils.

Group Five: Minority Children Who Use Non-conventional English with Power

All of the children in this group are Negro and all come from low socio-economic backgrounds. Because of their superiority in fluency, as well as their ability to phrase thought with complexity, these students resemble the children of Group One. Coming from minority family environments, they lack the necessary exposure to appropriate standard usage needed for its acquisition. Yet these ten children can be characterized as able, fluent speakers despite their non-standard speech.

Group Six: Children to Whom Standard English is a Second Dialect or a Foreign Language

Of the thirty children comprising this language group, 23 are Negro and three are Oriental. They are below average users of dependent clauses and also below average with respect to conventionality and elaboration. These children are accustomed to speaking some form of non-standard English at home. Their basic problem, using standard English in school, conflicts with the dialect that they normally hear and use at home. Unlike the children of Group Five who also speak with a dialect, these subjects have restricted syntactical repertoire, even with their own dialect.

Group Seven: Children Who are Incomprehensible Users of Non-conventional English

The children of this group are mainly from minority backgrounds. Their oral language is nearly incomprehensible to their teachers. Their problems with mazes are abnormally acute and their use of standard English is minimal. There is some possibility that damaged self-images may have interrupted their opportunities to learn how to use language effectively. Not only are they incoherent users of standard English; most of them are also unable to use their own non-standard variations with any facility.

Group Eight: Unconventional English Speakers Who Avoid the Use of Language

These students are non-talkers, but when they do speak, their speech is exceptionally unconventional. Their oral expression is very limited and the amount of information they express with labored speech is next to zero.

Univariate Analysis of Variance for Equality of Mean Values of the Eight Language Groups

Multivariate analysis of variance is the simple extension of the univariate analysis of variance to groups in which more than one dependent variable is observed. The differences and similarities between the two models are best illustrated by example. To simplify this comparison, consider the univariate analysis of variance for the null hypothesis of equal fluency measures at grade one. For the first grade statistics, the mean standardized scores for the eight groups of the study are as shown in Table 10. To determine whether the observed variation between the sample means is larger than could be expected on the basis of chance, a measure of variation called the mean square between the groups is com-

Table 10. Sample Statistics for the Fluency Measures at Grade One.

Group	Mean	Sample Size	Standard Deviation
1	60.4	27	5.3
2	53.9	51	5.8
3	41.9	24	6.5
4	51.5	37	6.3
5	63.3	10	6.8
6	45.1	30	5.9
7	47.7	16	7.4
8	31.8	16	6.9
Average	50.0	20.85*	6.2**

*This average is the harmonic mean of the sample sizes:

$$\overline{N} = \frac{G}{\sum_{g=1}^{G} \frac{1}{n_g}}$$

**This average is the pooled estimate of the standard deviation and in an analysis of variance model is called the square root of the mean square within groups;

$$S_p = \sqrt{MSW} = \sqrt{\sum_{g=1}^{G} \frac{(N_g - 1)S_g^2}{N - G}}$$

puted. This measure is defined by the following formula:

$$MSB = \frac{1}{G - 1} \sum_{g=1}^{G} n_g(\overline{X}_g - \overline{X})^2$$

where $\overline{X}_g$ = mean of the g-th group, $\overline{X}$ = grand mean of the entire set of data, n_g = number of students in the g-th group, and G = number of groups. For the observed means:

$$\begin{aligned} MSB &= \frac{1}{8-1}[27(60.4 - 50)^2 + 51(53.9 - 50)^2 + 24(41.9 - 50)^2 \\ &\quad + 37(51.5 - 50)^2 + 10(63.6 - 50)^2 + 30(45.1 - 50)^2 \\ &\quad + 16(47.7 - 50)^2 + 16(31.8 - 50)^2] \\ &= \frac{1}{7}[27(10.4)^2 + 51(3.9)^2 + 24(-8.1)^2 + 37(1.5)^2 \\ &\quad + 10(13.6)^2 + 30(-4.9)^2 + 16(-2.3)^2 + 16(-19.1)^2] \\ &= \frac{1}{7}[2920.32 + 775.71 + 1574.64 + 83.25 + 1849.60 \\ &\quad + 720.30 + 84.64 + 5836.96] \\ &= \frac{13845.42}{7} \\ &= 1977.92 \end{aligned}$$

To determine whether this number is indicative of true mean differences, it is compared to the variance within the group which is defined by the following formula:

$$MSW = \frac{1}{N - G} \sum_{g=1}^{G} (n_g - 1)S_g^2$$

where S_g^2 = variance of the g-th group. For the observed data:

$$\begin{aligned} MSW &= \frac{1}{211 - 8}[26(5.3)^2 + 50(5.8)^2 + 23(6.5)^2 \\ &\quad + 36(6.3)^2 + 9(6.8)^2 + 29(5.9)^2 + 15(7.4)^2 \\ &\quad + 15(6.9)^2] \\ &= 38.44 \end{aligned}$$

so that $S_p = \sqrt{38.44} = 6.2$

The comparison is made by way of a statistic called the F-ratio which is simply the ratio of MSB to MSW. For these data:

$$F = \frac{MSB}{MSW} = \frac{1977.92}{38.84} = 50.92$$

If the variation between the groups is larger than could be expected on the basis of chance, the F-ratio will be large, whereas if the variation is not unusual, the value of the F-ratio will be close to one in value. In this case, it is quite large so that one might suspect that the differences are larger than chance would predict. To determine whether the variation between the groups is larger than could be expected on the basis of the natural variability that exists within the groups, the computed F-ratio is referred to tabled values of F that are found in most elementary statistics texts. To enter these tables, two numbers need to be determined. These numbers are called the degrees of freedom of the mean square between and the degrees of freedom of the mean square within. The formulas for these numbers are given by $\nu_1 = G - 1 = 8 - 1 = 7$, and $\nu_2 = N - G = 211 - 8 = 203$, respectively. Since it is rarely possible to possess all of the information that exists concerning a variable, it is necessary to accept the possibility of making an error in concluding that a significant difference exists between the groups when none really does. This error is called a Type I error and its risk of occurrence is denoted by α where $0 \leq \alpha \leq 1$. Most researchers would like to maintain this risk of error as small as possible; to achieve this goal it is customary to set α equal to .05 or equal to .01. If the risk of error is set equal to .01, and if the F table is entered with $\nu_1 = 7$ and $\nu_2 = 203$, it is found that the variation between the means is larger than expected if $F > 2.64$. In this case, $F = 50.92$, so that the F-ratio is larger than expected. It is therefore concluded that a significant difference exists between the groups.

In correlation theory, it is possible to show that the square of the correlation coefficient, r, is a measure of explained variance. In the analysis of variance, there exists a similar measure defined by Hayes(14) which can be approximated by the following formula:

$$\omega^2 = \frac{SSB}{SST} = \frac{(G - 1)MSB}{(N - G)MSW + (G - 1)MSB}$$

For these data:

$$\hat{\omega}^2 = \frac{(8 - 1)(1977.92)}{(211 - 8)(38.84) + (8 - 1)(1977.92)}$$

$$= \frac{13845.44}{7884.52 + 13845.44}$$

$$= \frac{13845.44}{21729.96}$$

$$= 63.7\%$$

For behavioral data, this is an exceptionally high value of explained variance. It also indicates that the relative distances between the means of the eight samples is very large.

Once a tested null hypothesis has been rejected, the next task facing a researcher is to locate the differences in means that are larger than could be expected on the basis of chance factors alone. This identification is accomplished by the application of Scheffé's Theorem(15) to the observed means. In its simplest application, all pair-wise differences between the means are computed. If any difference exceeds

$C = \underline{S}(SE_d) = \sqrt{\nu_1 F_{\nu_1,\nu_2}(1 - \alpha)} \sqrt{\frac{MSW}{N_1} + \frac{MSW}{N_2}}$, it is concluded that a significant difference exists between the two groups being compared. $\underline{S}$ is called the Scheffé Coefficient, while C is called the critical value for the observed difference. For this model, the value of the Scheffé Coefficient is given by $\underline{S} = \sqrt{7(2.64)} = \sqrt{18.48} = 4.30$.

When the sample sizes are unequal, one can approximate the exact SE_d by $\overline{SE}_d = \sqrt{\frac{2}{\overline{N}} MSW}$ where $\overline{N}$ equals the harmonic mean of the sample sizes and which is defined by:

$$\overline{N} = \frac{G}{\frac{1}{n_1} + \frac{1}{n_2} + \ldots + \frac{1}{n_G}}$$

For these data:

$$\overline{N} = \frac{8}{\frac{1}{27} + \frac{1}{51} + \frac{1}{24} + \frac{1}{37} + \frac{1}{10} + \frac{1}{30} + \frac{1}{16} + \frac{1}{16}} = 20.85$$

For the observed means, the paired differences are shown in Table 11. With these figures and statistics:

$$\overline{SE}_d = \sqrt{\frac{2}{\overline{N}} MSW} = \sqrt{\frac{2}{20.85} (38.84)} = 1.93$$

and:

$$C = \underline{S}\overline{SE}_d = 4.30 \times 1.93 = 8.30$$

If each of the paired mean differences shown in Table 11 is compared to this value, it is seen that 19 pairs are statistically different. These significant differences are indicated with an asterisk in Table 11. It should be noticed that the numbers in Table 11 give rise to a symmetry about the main diagonal that differs only by algebraic sign. Such tables of paired mean differences will always have this symmetric

Table 11. Ordered Paired Differences of the Eight Means for Fluency Grade One.

Group	1	2	3	4	5	6	7	8
1	---	-6.5	-18.5*	-8.9*	2.9	-15.3*	-12.7*	-28.6*
2	6.5	---	-12.0*	-2.4	-9.3*	-8.8*	-6.2	-22.0*
3	18.5*	12.0*	---	9.6*	21.3*	3.2	5.8	-10.1*
4	8.9*	2.4	-9.6*	---	11.7*	-6.4	3.8	-19.7*
5	-2.9	9.3*	-21.3*	-11.7*	---	-18.1*	-15.5*	-31.4*
6	15.3*	8.8*	-3.2	6.4	18.1*	---	2.5	-13.3*
7	12.7*	6.2	-5.8	-3.8	15.5*	-2.5	---	-15.8*
8	28.6*	22.0*	10.1*	19.7*	31.4*	13.3*	15.8*	---

*Significant at $\alpha = .01$

form, and for this reason, the lower half of such tables will not be reported in the remainder of this narrative.

As can be seen, Group One is statistically different from all the other groups except Groups Two and Five. As is recalled, the students of these three groups were characterized as being quite fluent in speech and of considerable syntactic complexity in their general use of language. Note particularly that Group Five, the ten Negro subjects, are different from every other group except the highly superior language users, Group One. Once again, the ability of these ten Negro pupils to express themselves in complex forms marks them as a special and unusual group of subjects. As another interesting finding, it is seen that Group Eight is statistically different from all of the remaining seven groups generated by the clustering procedure, suggesting that these students, of limited capacity both intellectually and in terms of vitality, are extremely different from the students in the remaining seven language groups.

Multivariate Analysis of Variance for the Mean Vector of the Eight Language Groups

As might be expected, the univariate procedures described in the previous paragraphs are special cases of the multivariate procedures used in this study. In the univariate model, one tests the hypothesis that the variation existing for one single variable between the means of G different groups is larger than expected on the basis of chance. In the multivariate model, this notion is extended to a test that the variation existing between the corresponding means on each of p different variables is larger than could be expected on the basis of chance. In the univariate model, only one variable is being examined for mean differences while in the multivariate model all p variables are examined simultaneously for possible mean differences across the groups. Thus, if one considers the eight profiles shown graphically in Figures One through Eight, one would want to know whether the differences existing between the profiles, variable by variable, are larger than one could expect on the basis of chance. To answer this question, one again computes a statistic measuring the variation between the means of the profiles, and as was done in the univariate case, this observed measure of group variation is compared to the natural variation existing within the groups on all of the variables taken collectively. As might be expected, the statistical formulas one encounters in the multivariate model are extremely complex and not absolutely essential to an understanding of the discussion in this narrative. However, to aid the reader, it will be useful to introduce a minimal amount of notation, definitions, and formulas.

If the means of a profile are listed in the following manner:

$$(\overline{X}_F, \overline{X}_M, \overline{X}_D, \overline{X}_C, \overline{X}_E)$$

it is said that the profile has been represented in vector form. In the context of Table 9, the first row of numbers (60.5, 51.3, 62.0, 55.8, 61.9) represents the vector of means for the pupils of Group One at grades one, two, and three. In a more common mode of presentation, vectors are represented in a vertical fashion similar to the following:

$$X = \begin{bmatrix} 60.5 \\ 51.3 \\ 62.0 \\ 55.8 \\ 61.9 \end{bmatrix}$$

When the means are displayed in this manner, the vector is said to be a column vector. Column vectors are generally denoted by a capital X . If one wishes to denote a column vector in row form, one denotes the vector by X' . A column vector presented in row form is called the transposed vector of X .

Under this notation and set of definitions, it follows that rows of Table 9 represent eight vectors of means that summarize the profile data for each of the eight groups. Each row represents a transposed vector of means. Such a two-dimensional array of numbers is called a matrix. When the matrix consists of transposed vectors, the matrix is called a transposed matrix. Thus, if X represents a matrix with r rows and c columns, then X' represents a matrix with c rows and r columns.

Matrices need not be restricted to a display of means. Any sort of numbers can be described in a matrix. Three very important matrices which will be employed in the analysis of the Loban data are T^*, S^*, and B^*,

$$T^* = \begin{bmatrix} S_1^2 & r_{12}S_1S_2 & r_{13}S_1S_3 & \cdots & r_{1p}S_1S_p \\ r_{21}S_2S_1 & S_2^2 & r_{23}S_2S_3 & \cdots & r_{2p}S_2S_p \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{p1}S_pS_1 & r_{p2}S_pS_2 & r_{p3}S_pS_3 & \cdots & S_p^2 \end{bmatrix}$$

is called the Variance-Covariance Matrix of the data. S_1, S_2, ..., are the standard deviations of the individual variables, while r_{12}, r_{13}, ..., represent the correlation coefficient between variables one and two, between one and three, ... When $S_1 = S_2 = \ldots = S_p = 1$, the Variance-Covariance Matrix reduces to:

$$R^* = \begin{bmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1p} \\ r_{21} & 1 & r_{23} & \cdots & r_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{p1} & r_{p2} & r_{p3} & \cdots & 1 \end{bmatrix}$$

In this form, the Variance-Covariance Matrix is called the Correlation Matrix. The Correlation Matrix for the Loban data at the early grades is as presented in Table 1.

The S^* matrix is very similar to the T^* matrix except that in place of the total variable variances, mean square within or pooled estimates of the variances are substituted. In addition, in place of variable correlation coefficients, an average correlation coefficient determined from each of the groups and so weighted to give the best possible estimate of the within sample correlation coefficients are substituted. With these substitutions, the S^* matrix is defined by:

$$\mathcal{S}^* = \begin{bmatrix} MSW_1 & \bar{r}_{12}\sqrt{MSW_1}\sqrt{MSW_2} & \cdots & \bar{r}_{1p}\sqrt{MSW_1}\sqrt{MSW_p} \\ \bar{r}_{21}\sqrt{MSW_2}\sqrt{MSW_1} & MSW_2 & \cdots & \bar{r}_{2p}\sqrt{MSW_2}\sqrt{MSW_p} \\ \vdots & \vdots & \cdots & \vdots \\ \bar{r}_{p1}\sqrt{MSW_p}\sqrt{MSW_1} & \bar{r}_{p2}\sqrt{MSW_p}\sqrt{MSW_2} & \cdots & MSW_p \end{bmatrix}$$

Finally, the $\mathcal{B}^*$ matrix consists of elements similar to the MSB of the univariate analysis of variance model. The elements on the main diagonal are given by the mean square between for each of the p variables, MSB_1, MSB_2, ..., MSB_p, but for the off diagonal elements, the following algebraic quantities appear:

$$MS(B_p, B_{p'}) = \frac{1}{G - 1} \sum_{g=1}^{G} n_g(\bar{x}_{gp} - \bar{x}_p)(\bar{x}_{gp'} - \bar{x}_{p'})$$

With this notation, the matrix $\mathcal{B}^*$ is defined as:

$$\mathcal{B}^* \begin{bmatrix} MSB_1 & MS(B_1,B_2) & \cdots & MS(B_1,B_p) \\ MS(B_2,B_1) & MSB_2 & \cdots & MS(B_2,B_p) \\ \vdots & \vdots & \cdots & \vdots \\ MS(B_p,B_1) & MS(B_p,B_2) & \cdots & MSB_p \end{bmatrix}$$

When p = 1:

$$\mathcal{B}^* = [MSB] \qquad \text{and} \qquad \mathcal{S}^* = [MSW]$$

so that the univariate F-ratio in matrix notation reduced to:

$$F = \frac{MSB}{MSW} = \frac{\mathcal{B}^*}{\mathcal{S}^*}$$

As this suggests, the univariate model can be subsumed under the p variable model. As an aside, it should be noted that:

$$\frac{B^*}{S^*} = \frac{\frac{1}{G-1}\sum_{g=1}^{G} n_G(\bar{x}_g - \bar{x})^2}{\frac{1}{N-G}\sum_{g=1}^{G}\sum_{i=1}^{N_g}(x_{ig} - \bar{x}_g)^2} = \frac{\frac{1}{G-1}[SSB]}{\frac{1}{N-G}[SSW]} = \frac{N-G}{G-1}\left[\frac{SSB}{SSW}\right]$$

where SSB = sum of squares between groups, and SSW = sum of squares within groups. Since $\frac{N-G}{G-1}$ is a numerical constant, the test of equal mean values could be performed by examining only $\frac{SSB}{SSW}$. In the multivariate model this practice is adhered to because it simplifies the arithmetic. Thus, for the p variable model, the basic matrices are:

$$\mathbf{B} = [SSB] = \begin{bmatrix} SSB_1 & SS(B_1,B_2) & \cdots & SS(B_1,B_p) \\ SS(B_2,B_1) & SSB_2 & \cdots & SS(B_2,B_p) \\ \vdots & \vdots & \vdots & \vdots \\ SS(B_p,B_1) & SS(B_p,B_2) & \cdots & SSB_p \end{bmatrix}$$

and:

$$\mathbf{W} = [SSW] = \begin{bmatrix} SSW_1 & SS(W_1,W_2) & \cdots & SS(W_1,W_p) \\ SS(W_2,W_1) & SSW_2 & \cdots & SS(W_2,W_p) \\ \vdots & \vdots & \vdots & \vdots \\ SS(W_p,W_1) & SS(W_p,W_2) & \cdots & SSW_p \end{bmatrix}$$

Since $\mathbf{B}$ and $\mathbf{W}$ are matrices with p rows and p columns, it is possible to determine a number associated with each matrix. These numbers are called the determinant of the matrix and are denoted by $|B|$ and $|W|$. From these numbers one defines the following ratio:

$$\Lambda = \frac{|W|}{|B + W|} = \frac{|W|}{|T|}$$

This number is called Wilk's Criteria(10) and is a multivariate analog to the univariate F-ratio. It is also the statistic used in the Friedmann-Rubin clustering procedure employed to define the eight language groups of this study.

When the between groups measure is small, it is seen that Λ is close to one in numerical value, while if the between component is large, Λ will tend to be close to zero in numerical value. Thus, small values of Λ are compatible with the rejection of the hypothesis of similar profiles. Unfortunately, the probability distribution of Λ is quite complex and not tabled. Fortunately, there is a simple relationship between Λ and the Chi-square distribution. This relationship is given by:

$$\chi^2 = -[N - G - \frac{1}{2}(p - G + 2)] \log_e \Lambda$$

This variable has a χ^2 distribution with $\nu = (G - 1)p$ degrees of freedom. When N is large, $(G - 1)p\chi^2$ is approximately distributed as $F_{(G-1)p,\nu_2}$ where ν_2 is very large. Box(11) has found closer approximations for this variable. This Box approximation is used in this narrative.

Another procedure which can be used to test for identical profiles and is also used in this narrative, is to find the solution for λ of the determinantal equation:

$$|B - \lambda W| = 0$$

that gives the maximum value of λ. For this procedure one compares λ^{max} to tabled values in a manner similar to that used for the F-table of the univariate model. Fortunately, the probability distribution of λ^{max} has been studied by Heck$^{(16)}$ by means of the following transformation:

$$\theta = \frac{\lambda}{1 + \lambda}$$

The parameters of the sampling distribution of θ are:

$$s = \min(G - 1, p)$$

$$m = \frac{|G - 1 - p| - 1}{2}$$

and:

$$n = \frac{N - G - p - 1}{2}$$

Under the Heck procedure, the hypothesis of equal profiles is rejected if $\theta > x_\alpha$, where the probability of a Type I error is set equal to α. If the hypothesis is rejected, then post hoc comparisons between the means within each of the variables is conducted in the same manner as that employed for the univariate case except that for $\underline{S}$, one uses:

$$\underline{S} = \sqrt{\nu_2 \frac{x_\alpha}{1 - x_\alpha}}$$

where ν_2 refers to the number of degrees of freedom for MSW_p. For the p-th variable, one has:

$$(\bar{x}_{p_1} - \bar{x}_{p_2}) - \underline{S}\sqrt{\frac{MSW_p}{N_1} + \frac{MSW_p}{N_2}} < \mu_{p_1} - \mu_{p_2} < (\bar{x}_{p_1} - \bar{x}_{p_2}) + \underline{S}\sqrt{\frac{MSW_p}{N_1} + \frac{MSW_p}{N_2}}$$

If zero is included in the interval, the null hypothesis of no difference is not rejected. The procedure was developed by Roy and Bose$^{(17)}$.

Another way to reach the same decisions is to compute:

$$C_p = \underline{S}\sqrt{\frac{MSW_p}{N_1} + \frac{MSW_p}{N_2}}$$

If $(\overline{x}_{p_1} - \overline{x}_{p_2}) < C_p$, then the null hypothesis of no difference is not rejected; but if $(\overline{x}_{p_1} - \overline{x}_{p_2}) > C_p$, it is concluded that a significant difference between the means exists.

Multivariate Analysis of the Average T Scores for the First, Second, and Third Grade Data

The basic statistics for the eight language groups are summarized in Table 12. The average scores were first shown in Table 8 and at that time were used to define the eight language groups of the study. They were also used to prepare the graphic profiles shown in Figures 1 through 8. Graphic representations of the observed distributions are shown in Figures 9 through 13.

As can be seen, the subjects who comprise the individual language groups are quite homogeneous with respect to language since the standard deviations of the individual groups are quite small when compared to the unconditional standard deviation of the basic T variables which are all equal to 10 in numerical value. Only two of the eight language groups show any unusually large variability with respect to any of the five language variables. These are Groups Seven and Eight. As noted earlier, subjects in Group Seven have particular problems with mazes in that their average standardized score is equal to 36.3. In addition, the large standard deviation for this variable of 10.5 standardized units suggests that the problem is exceptionally acute for some of the children in this group. These subjects speak a nonstandard English and belong to families who do not instill a tradition of striving to succeed in school. Part of the large

Table 12. Basic Statistics for the Standardized Language Variables of the Eight Language Groups at Grades One, Two, and Three.

		Group								
Statistic	Variable	1	2	3	4	5	6	7	8	Average
Averages	Fluency	60.5	53.5	41.2	52.4	60.2	45.2	49.1	32.4	50
	Mazes	51.3	53.2	57.3	48.0	38.6	48.3	36.3	56.0	50
	Dependent Clauses	62.0	51.6	44.0	51.5	60.0	43.6	46.2	39.6	50
	Conventionality	55.8	56.3	55.4	50.3	46.2	42.8	39.7	37.7	50
	Elaboration Index	61.9	52.3	42.6	52.1	58.5	44.9	47.4	35.8	50
Standard Deviation	Fluency	4.1	4.5	5.3	4.3	5.6	3.6	6.6	5.7	4.7**
	Mazes	5.0	5.6	4.9	5.7	4.4	6.7	10.5	7.7	6.3**
	Dependent Clauses	7.5	4.1	4.0	4.3	5.4	2.8	3.5	3.0	4.5**
	Conventionality	2.8	2.8	2.8	6.0	4.9	7.9	7.1	13.5	6.2**
	Elaboration Index	4.9	3.3	3.0	4.4	4.4	3.4	4.8	5.1	4.0**
Sample Size		27	51	24	37	10	30	16	16	20.85*

*Harmonic Mean

**Within sample estimate of the standard deviation

Figure 9. Histograms of the Fluency Measures for the Eight Language Groups.

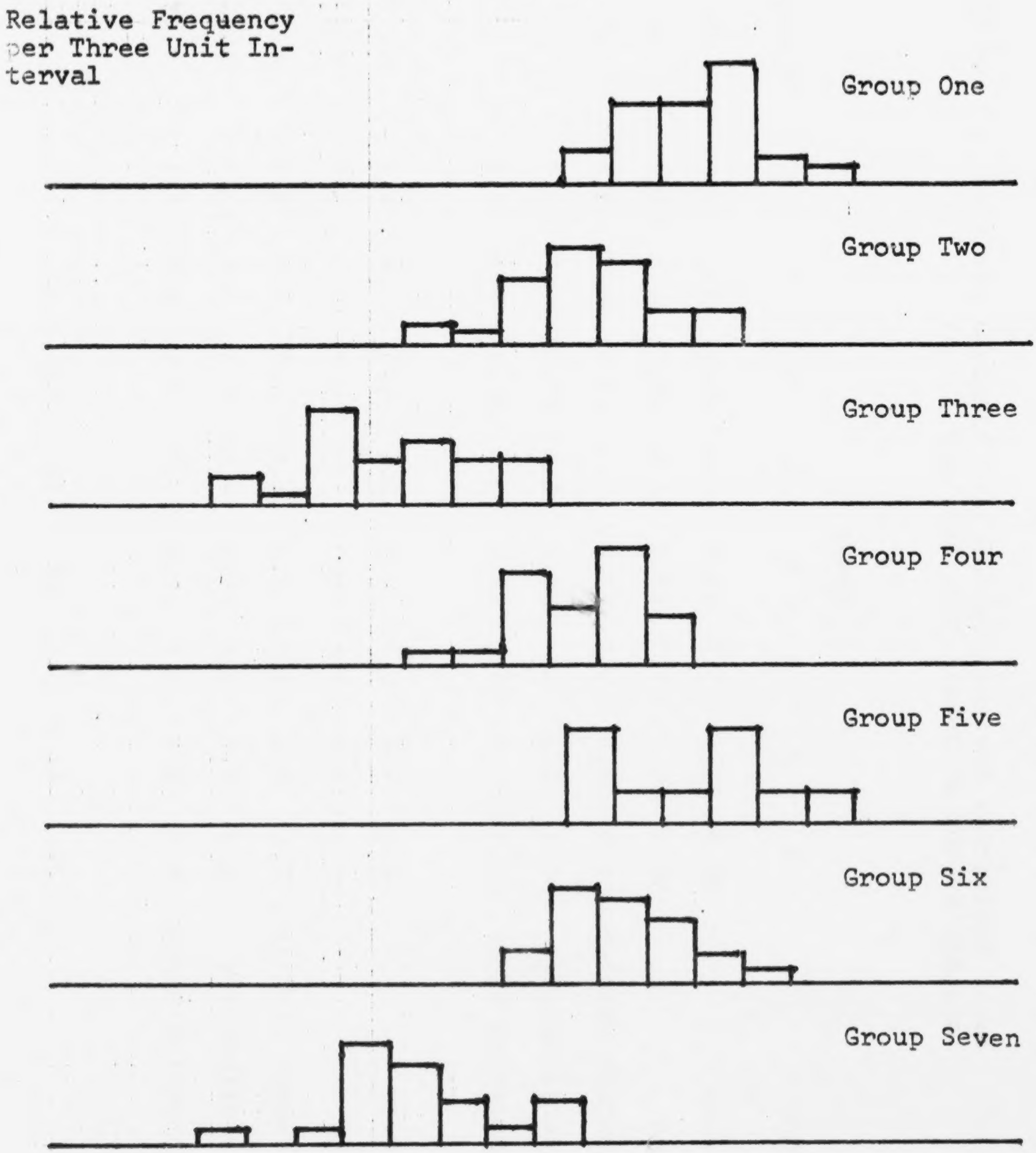


Figure 10. Histograms of the "Maze" Measures for the Eight Language Groups.

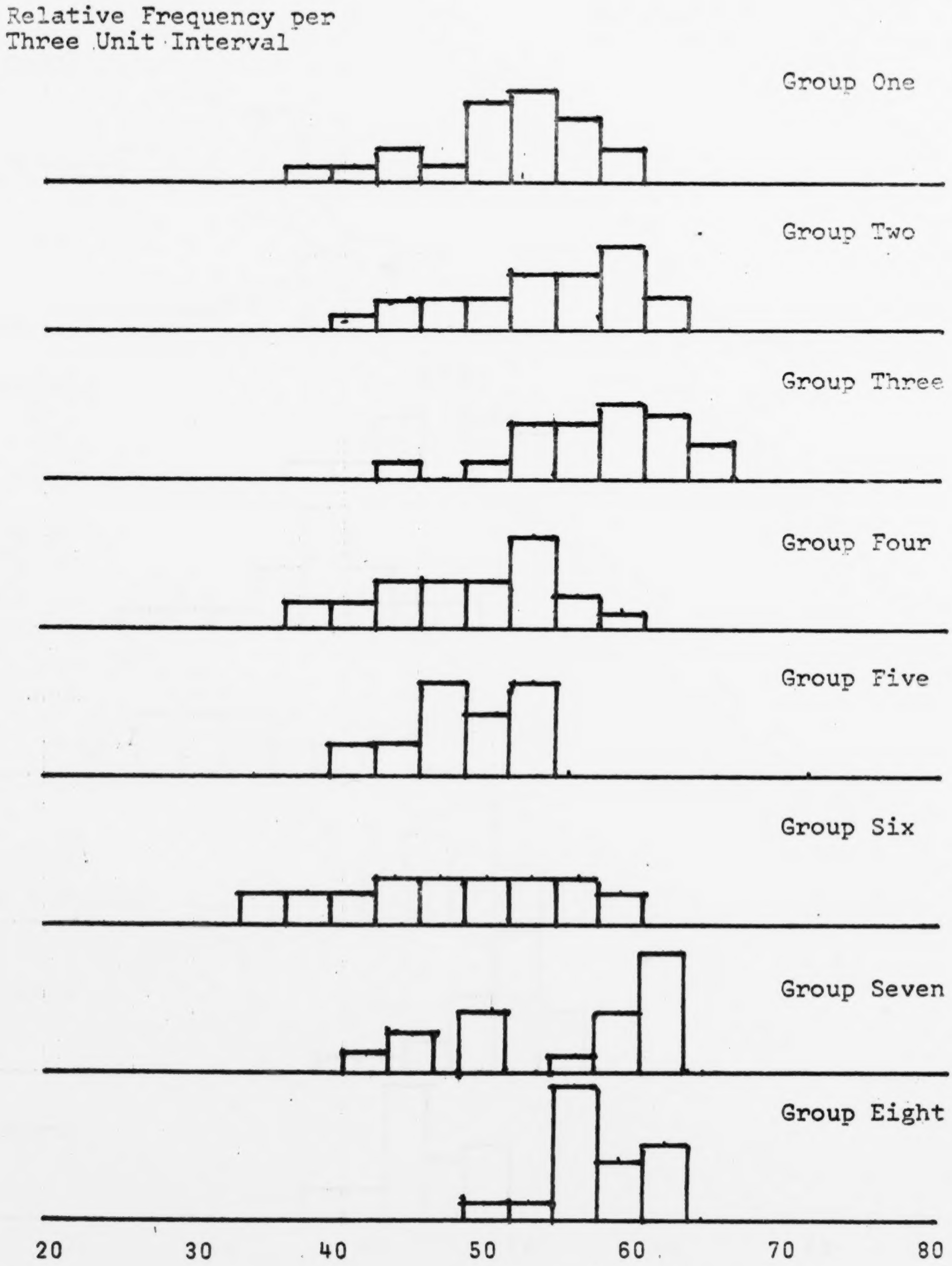


Figure 11. Histograms of the Dependent Clause Measurement for the Eight Language Groups.

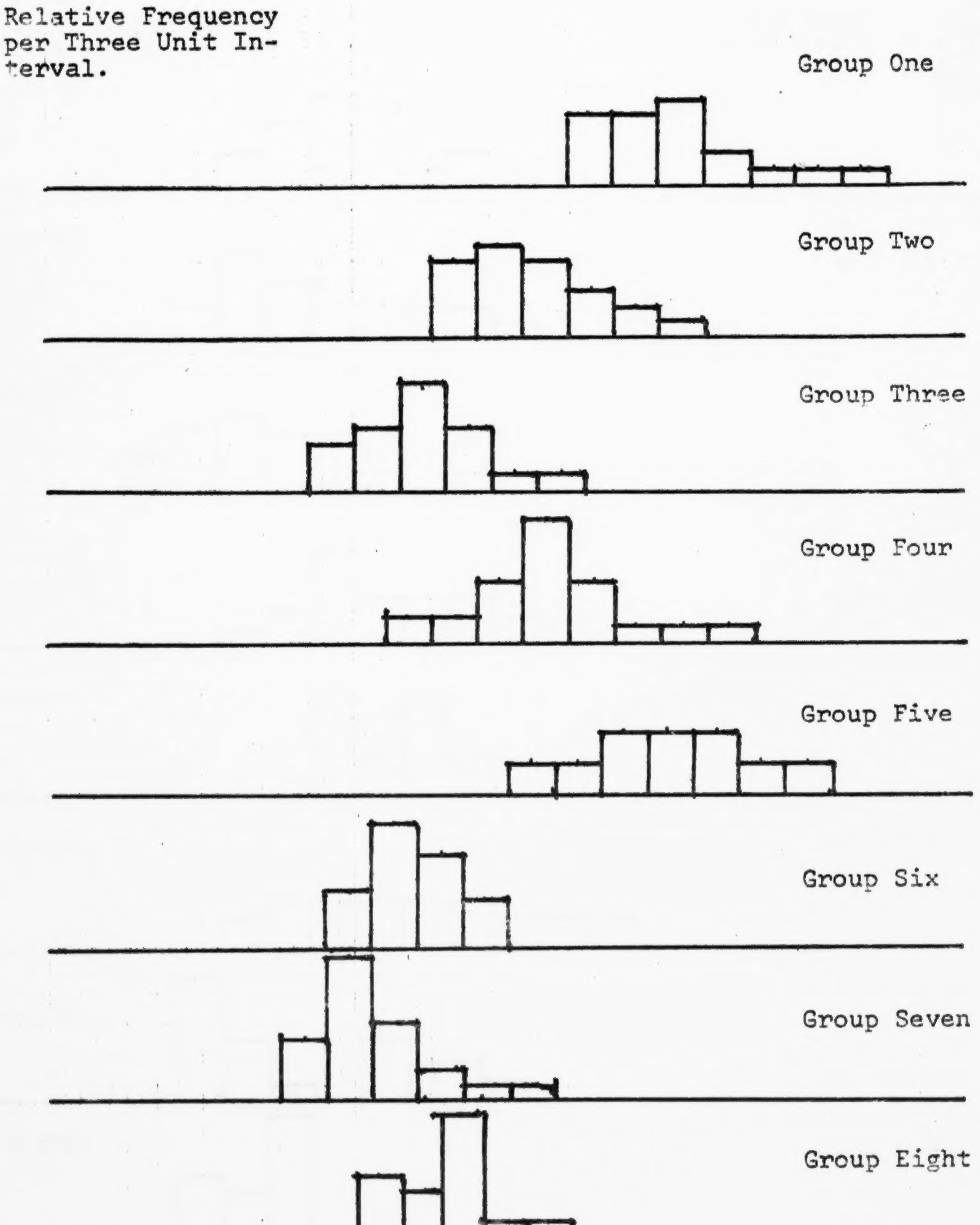


Figure 12. Histograms of the Conventionality Measures for the Eight Language Groups.

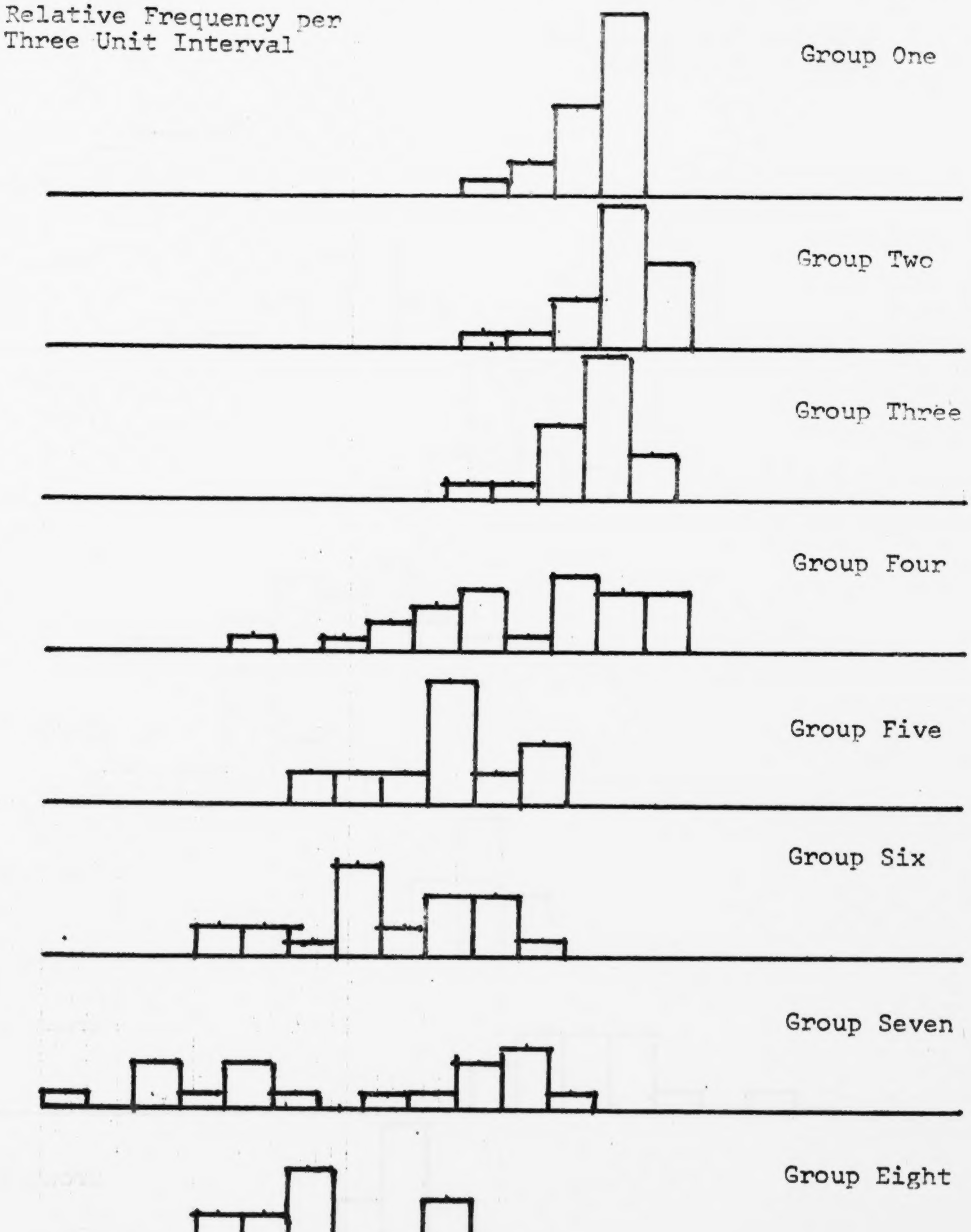
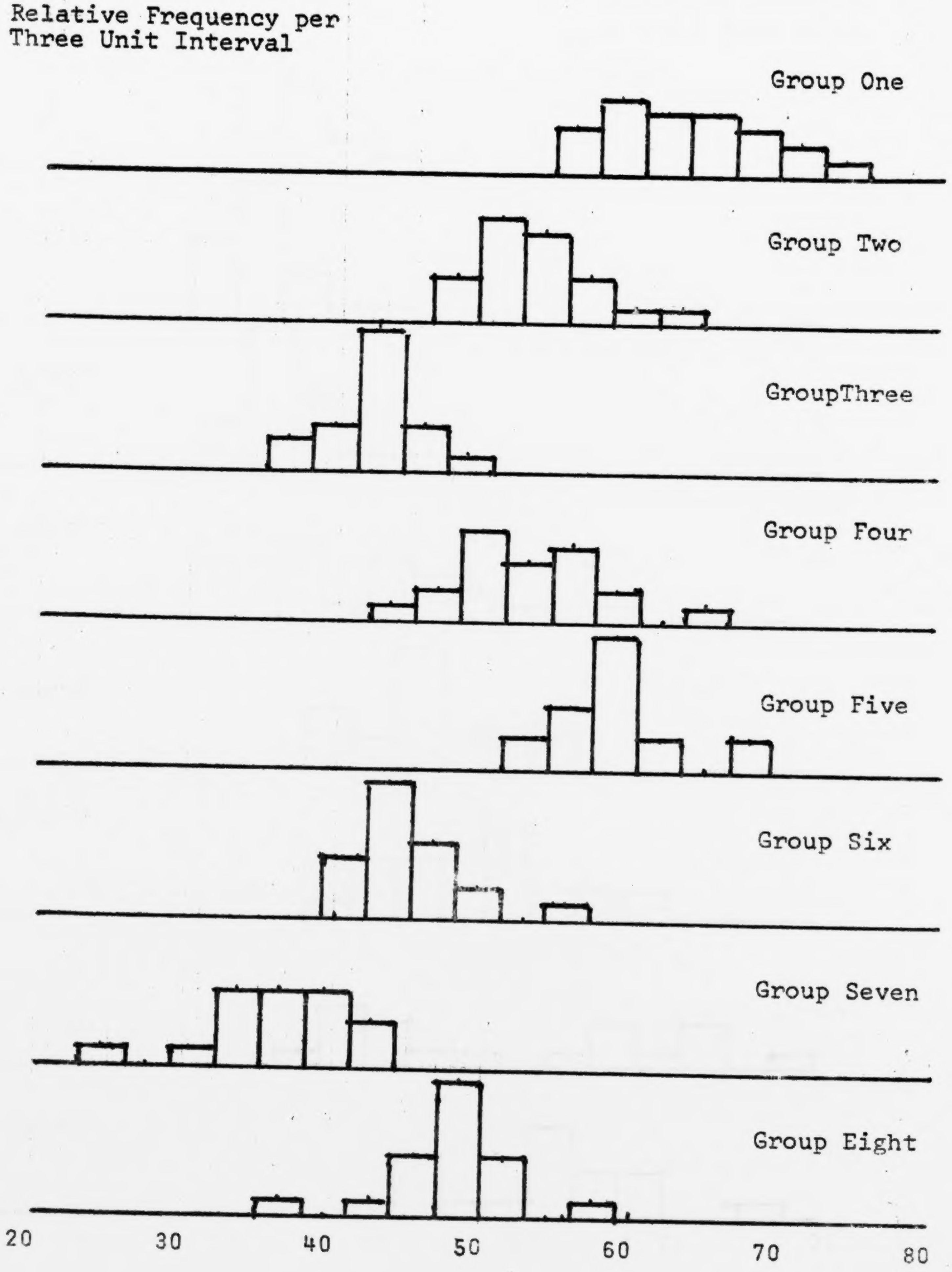


Figure 13. Histograms of the Elaboration Index for the Eight Language Groups



variance of this variable for these children might be related to these characteristics. As is recalled, students in Group Eight demonstrated the most problems with conventionality of speech with a mean standardized score of only 37.7. For these students the standard deviation of the conventionality measure is exceptionally large at 13.5 standard score points. Since it was noted that two different groups of students seemed to be combined to make one group by the Friedmann-Rubin program, it is quite possible that the large variance on this variable reflects this possible combining of two different groups of students into one group.

Examination of the matrix of standard deviation suggests that the variance within the eight language groups is quite uniform across the groups, variable by variable, except for the measures of conventionality and freedom from mazes. For language Groups One, Two, and Three, the conventionality scores are very homogeneous. The mean standard deviation is only 2.8 standard units in length. As recalled, these three groups contain students who are either conventional in speech or who tend to use language in a powerful manner. The greater variability in conventionality is noted for those groups consisting of students who are less able with language and are members of Groups Six, Seven, and Eight. Many of these students are from minority families with low socio-economic status.

The correlation matrix for the average standardized language variables for grades one, two, and three is as shown in Table 13. As can be seen, the average correlations within

Table 13. The Within Correlation Matrix-R for the Standardized Language Variables at Grades One, Two, and Three.

Variable	Fluency	Mazes	Dependent Clauses	Conven tionality	Elaboration Index
Fluency	1.00	-.02	-.04	-.07	.26
Mazes		1.00	.13	.15	.21
Dependent Clauses			1.00	.04	.41
Conventionality				1.00	-.07
Elaboration Index					1.00

a group are quite low and close to zero in numerical value, suggesting that within a language group the five language variables are providing unique measures on five different language characteristics. This is a somewhat different interpretation than that provided for the data of Tables 1 and 4 in which the large correlation measures were used to advance the argument of redundancy of information to justify the use of principal components. In those tables, the differences that exist between the groups on the various language variables have not been partialed out so that the total correlations are much larger than one would find within the language groups. A visual explanation for this paradox is suggested in Figure 14 which shows eight subgroups in which the correlations are all equal to zero, but for which the total correlation between the two variables is very high when considered across the total range of the two variables.

The B and W matrices for the observed data are shown in Table 14. With these matrices, the value of the F-ratio, as approximated by the Box procedure$^{(11)}$, is given by $F = 26.56$. In this case $\nu_1 = (G - 1)p = (8 - 1)(5) = 35$, and $\nu_2 = 839 \sim \infty$. With $\alpha = .05$, the hypothesis of equal mean vectors or identical language profiles should be rejected if $F > F_{35,\infty}(.95) = 1.43$. Since $F = 26.56 > 1.43$, the hypothesis of equal mean vectors is rejected. It is now known that the language profiles are statistically different. This should come as no surprise since the eight language groups were so constructed by the Friedmann-Rubin program so as to give maximum differences between groups with respect to the average principal component

Figure 14. Illustration of the Model in Which the Correlation Within Groups is Zero, but for Which the Correlation Across Groups is Very High.

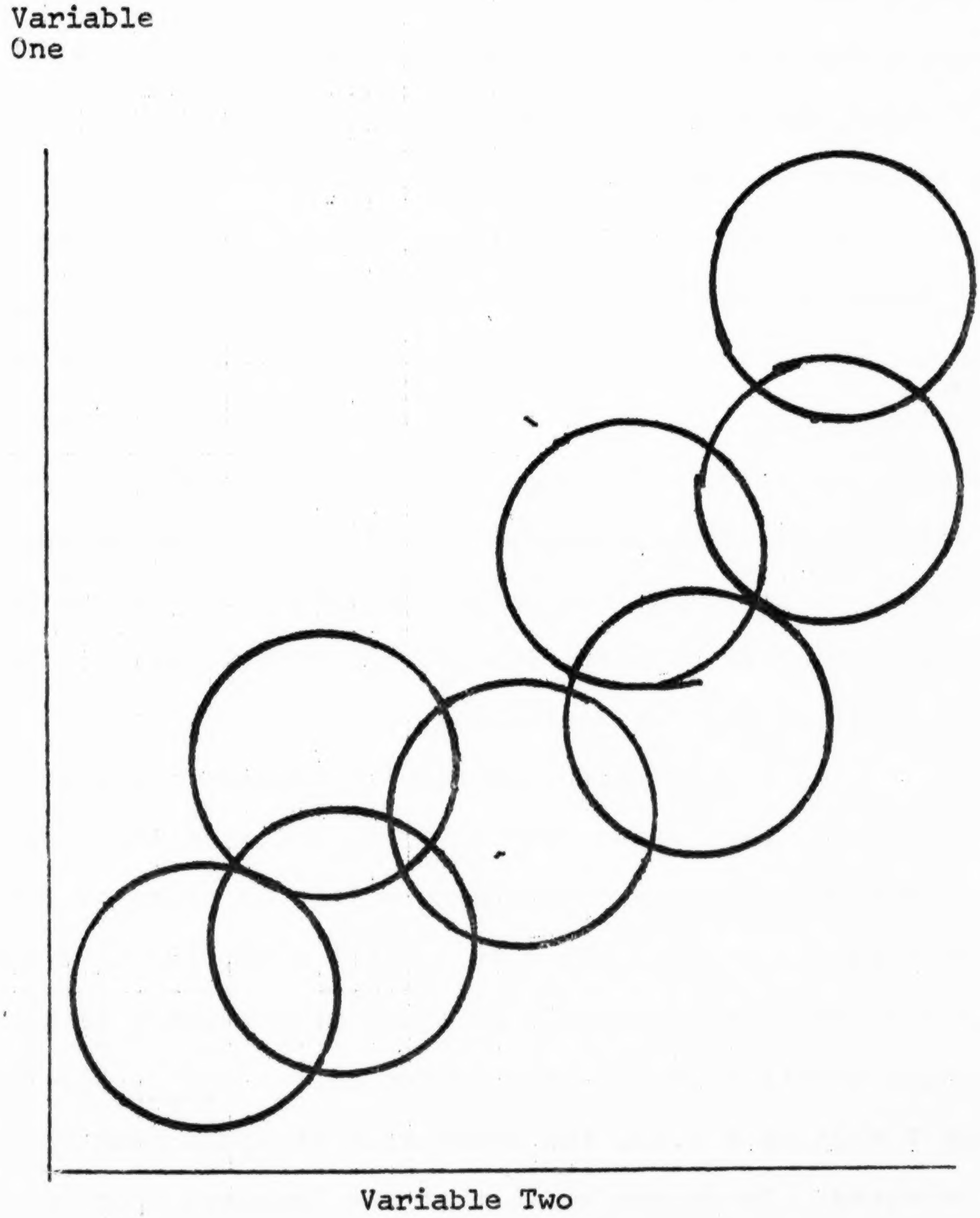

Variable One
Variable Two

Table 14. The B and W Matrices for the First, Second, and Third Grade Average T Scores.

Matrix	Variable	Fluency	Mazes	Dependent Clauses	Conventionality	Elaboration Index
B	Fluency	1764.9	462.7	1430.0	842.6	1594.1
	Mazes		991.2	213.7	574.0	314.1
	Depen. Clauses			1308.3	-762.2	1367.6
	Conventionality				1354.6	-811.3
	Elabor. Index					1485.6
W	Fluency	4547.2	101.5	-166.5	412.1	996.7
	Mazes		7937.3	-724.7	1193.6	-1065.8
	Depen. Clauses			4141.2	-227.4	1508.3
	Conventionality				7571.9	-371.5
	Elabor. Index					3329.2

scores on two dimensions. As is recalled, the two dimensions created by the principal component analysis were constructed, in part, from these five language variables.

For completeness, the univariate analyses of variances for the five language variables are shown in Table 15. As can be seen, all five measures of explained variance exceed 40 percent, indicating that large differences exist between the groups on each of the five language variables. If the individual variable hypotheses are controlled with $\alpha = .01$, then the hypothesis of identical mean values should be rejected if $F > F_{G-1,N-G}(.99) = F_{7,203}(.99) = 2.08$. Since all F-ratios exceed this value, the hypotheses of identical mean values are rejected for each variable. The reason .01 is used for testing each individual variable is that it controls the maximum probability of a Type I error for the complete set of hypotheses equal to $p\alpha = 5(.01) = .05$, the value used for the multivariate test of identical profiles.

For these data, the numerical value of the Roy criteria is given by $\theta = .8669$. For this measure of identical language profiles, the parameters of the sampling distributions are given by:

$$s = \min(G - 1, p) = \min(7, 5) = 5$$

$$m = \frac{|G - 1 - p| - 1}{2} = \frac{|8 - 1 - 5| - 1}{2} = \frac{2 - 1}{2} = \frac{1}{2} = .5$$

$$n = \frac{N - G - p - 1}{2} = \frac{211 - 8 - 5 - 1}{2} = 98.5$$

For $\alpha = .05$, the critical value is given by $x_\alpha = .11$. Since $\theta = .8669 > .11$, the hypothesis of identical profiles is re-

Table 15. Univariate Analysis of Variance for the Five Variables at Grades One, Two, and Three.

Variable	Source of Variance	d/f	Mean Square	F-ratio	$\hat{\omega}^2$	Decision*
Fluency	Between Groups	7	1764.9	78.6	.73	Reject
	Within Groups	203	22.5			
	Total	210				
Mazes	Between Groups	7	991.2	25.3	.47	Reject
	Within Groups	203	39.2			
	Total	210				
Depen. Clauses	Between Groups	7	1308.3	64.1	.41	Reject
	Within Groups	203	20.4			
	Total	210				
Conven.	Between Groups	7	1354.6	36.3	.56	Reject
	Within Groups	203	37.3			
	Total	210				
Elabor. Index	Between Groups	7	1485.6	90.6	.76	Reject
	Within Groups	203	16.4			
	Total	210				

*Hypothesis is rejected if $F > F_{7,203}(.99) = 2.08$

jected. This agrees with the decision made on the basis of the Box form of the F-test.

Now that the hypothesis of identical profiles has been rejected, the next task is to locate possible sources for the rejection and to identify the significant differences between the pairwise means for the five language variables. For this analysis, Scheffé intervals with $\underline{S} = \sqrt{\nu_2 \frac{x_\alpha}{1 - x_\alpha}}$ are employed. For these data:

$$\underline{S} = \sqrt{203 \frac{(.11)}{(1 - .11)}} = \sqrt{25.09} = 5.01$$

so that the approximate confidence intervals of interest using the harmonic mean on the sample sizes is given by:

$$\mu_{gp} - \mu_{g'p} = (\bar{x}_{gp} - \bar{x}_{g'p}) \pm 5.01\sqrt{\frac{MSW_p}{\bar{N}} + \frac{MSW_p}{\bar{N}}}$$

$$= (\bar{x}_{gp} - \bar{x}_{g'p}) \pm C_p$$

For the five language variables, the values of C_p are as shown in Table 16.

Scheffé-type Analysis of the Differences Between the Mean Vectors of the Eight Language Groups at Grades One, Two, and Three.

The pair-wise differences in mean values for the five language variables across the eight language groups are presented in Table 17. In the first matrix of the table of mean differences, Group One is compared to each of the remaining seven language groups. As can be seen, Group One, consisting of subjects who use language with clarity, ease, and power, is statistically different from each of the remaining language groups on at least one language variable. With respect to

Table 16. Computations Leading to Values of C_p; the Critical Value for the Pairwise Differences in Means for the Five Basic Language Variables at Grades One, Two, and Three.

Variable	MSW_p	$\frac{2}{N} MSW_p$	C_p^2	C_p
Fluency	22.4533	2.1537	54.04	7.35
Mazes	39.1768	3.7578	94.28	9.71
Dependent Clauses	20.4074	1.9575	49.11	7.00
Conventionality	37.3351	3.5812	89.85	9.48
Elaboration Index	16.3998	1.5731	39.47	6.28

Table 17. Variable by Variable Pairwise Comparisons Between the Profiles of the Eight Language Groups.

Group	1	2	3	4	5	6	7	8
Group One Versus the Seven Remaining Groups								
Fluency		7.0	19.3*	8.1*	.3	15.3*	11.4*	28.1*
Mazes		2.0	6.2	-3.1	-12.6*	-2.8	-14.9*	4.8
Dep.Cl.		10.4*	18.0*	10.5*	2.0	18.4*	15.8*	22.4*
Conven.		.4	-.4	-5.6	-9.7*	-13.0*	-16.2*	-18.2*
Elab.Ind.		9.6*	19.3*	9.8*	3.4	17.0*	14.5*	26.1*
Group Two Versus the Seven Remaining Groups								
Fluency			12.3*	1.1	-6.7	8.3*	4.3	21.0*
Mazes			4.2	-5.1	-14.5*	-4.8	-16.9*	2.8
Dep.Cl.			7.6*	.1	-8.3*	8.1*	5.4	12.0*
Conven.			-.8	-6.0	-10.1*	-13.4*	-16.6*	-18.6*
Elab.Ind.			9.7*	.2	-6.2	7.4*	4.9	16.5*
Group Three Versus the Seven Remaining Groups								
Fluency				-11.2*	-19.0*	-4.0	-7.9*	8.7*
Mazes				-9.2	-18.7*	-9.0	-21.0*	-1.4
Dep.Cl.				-7.5*	-15.9*	.5	-2.2	4.4
Conven.				-5.2	-9.1	-12.6*	-15.8*	-17.8*
Elab.Ind.				-9.5	-15.9*	-2.3	-4.8	6.8*

Table 17. (Continued)

Group	1	2	3	4	5	6	7	8
Group Four Versus the Seven Remaining Groups								
Fluency					-7.8*	7.2	3.3	19.9*
Mazes					-9.5	.3	-11.8*	7.9
Dep.Cl.					-8.5*	7.9*	5.2	11.9*
Conven.					-4.1	-7.5	-10.6*	-12.6*
Elab.Ind.					-6.4*	7.2*	4.7	16.2*
Group Five Versus the Seven Remaining Groups								
Fluency						15.0*	11.0*	27.7*
Mazes						9.7*	-2.3	17.4*
Dep.Cl.						16.4*	13.7*	20.3*
Conven.						-3.4	-6.6	-8.5
Elab.Ind.						13.6*	11.1*	22.7*
Group Six Versus the Seven Remaining Groups								
Fluency							-3.9	12.7*
Mazes							-12.0*	7.6
Dep.Cl.							-2.7	3.9
Conven.							-3.2	-5.2
Elab.Ind.							-2.5	9.1*
Group Seven Versus the Seven Remaining Groups								
Fluency								16.7*
Mazes								19.7*
Dep.Cl.								6.6
Conven.								-2.0
Elab.Ind.								11.6*

fluency, dependent clauses and elaboration index, the subjects in Group One are most like those in Group Five. As is recalled, all 10 children comprising Group Five are Negro. They are also characterized as talkative children who come from culturally deprived backgrounds and who use non-standard speech patterns. The similarity of their language profile to the profile of children in Group One--characterized as children with broad-based vocabularies, superior, fluent, conventional English, and effective complex syntactical structures--is striking.

The group of subjects revealing the greatest deviation from the Group One subjects with respect to mazes was Group Seven. This latter group consisted of children whose use of English was non-standard and difficult to understand.

Children in Group Two were similar to Group One students with respect to fluency, freedom from mazes, and conventionality, but their use of dependent clauses and language elaboration was considerably unlike that of children in Group One. As is recalled, the children in Group Two differ qualitatively from those in Group One in that they fail to put the power into their speech that is characteristic of the children in Group One. Like the children of Group One, they experience no difficulties in fluency or conventionality. They are impressive in ability to communicate, but their use of language lacks color, texture, and imagery.

The children in Groups Three and Four differ from those in Group One with respect to fluency, use of dependent clauses, and elaboration of thought. The children in these two groups were either average typical middle class users of American English or else they were somewhat shy in verbal expression of their thoughts, feelings, and needs.

Finally, the children in Group One are most unlike those in Groups Six, Seven, and Eight with respect to language usage. All three of the latter groups contain a large proportion of subjects who do not use standard English, and/or do not hear standard English in their home setting. For some of these children, English is a foreign language, and for some, school is frustrating and non-rewarding. In some cases, teachers do not know how to reach these children, or if they can, they do not know what are the best methods or procedures that would help them speak appropriate English.

According to the vector of mean differences, (11.4*, -14.9*, 15.8*, -16.2*, 14.5), the students of Group Seven are the complete opposite of those in Group One with respect to use of language. As is recalled, the Group Seven children come from low SES families, have a median IQ of 80, and on teacher ratings of oral language have a low median rating of 2.30. Unlike the subjects of Group One, they are exceptionally prone to apathy in the school environment and are ill at ease in most classrooms. In addition, it seems reasonable to assume that their poor self-images undoubtedly influence their poor success with language.

In the second matrix of mean difference, the subjects in Group Two prove to be non-statistically different from the students in Group Four. The vector of mean differences (1.1, -5.1, .1, -6.0, .2) for these groups shows none of the paired variable differences as statistically different. Even though the differences between the overall profiles of language are quite similar, there are a significant number of qualitative differences between the groups which suggests that the groups are different from one another on dimensions other than that of language. For example, the subjects in Group Two have a median IQ of 114 while those in Group Four have a median IQ of 99. On the Kuhlmann-Anderson IQ scale this represents a one standard deviation difference in averages. Their teacher ratings are 3.75 and 3.14, also suggesting a qualitative difference between the two groups.

Except for the similarities between Group Two and Group Four subjects, those in Groups Two differ from the remaining groups on at least two language variables. Group Two differs from Group Three with respect to fluency, dependent clauses, and elaboration index, while it differs from Group Seven with respect to mazes and conventionality. In many respects, the subjects of Group Three are like those in Group Four with one major difference: Those in Group Three appear to have damaged self-images and lack of confidence with respect to speech and life in general. The profile differences in the fluency of their language, use of dependent clauses, and elaboration of ideas most likely reflects their non-use of speech. When they do use language, they speak with a major maze problem and in a conventional mode similar to that employed by students in Group Two. As was noted earlier, many of these children come from home backgrounds where parents demonstrated considerable concern about educational success. This leads one to speculate that if these parents were not so vocal about academic achievement in the presence of their children, these children might be similar to those in Group Two.

For the third matrix of mean differences, it is seen that the subjects in Group Three are most like those in Group Six with respect to their language profiles. However, according to their mean difference vector (-4.0, -9.0, .5, -12.6*, -2.2) the differences between these groups with respect to conventionality is quite large. As is recalled, Group Three subjects tend to be non-talkers because of personality inadequacies and timidity. Those of Group Six also tend to be non-talkers, not because of shyness but because English tends to be used by them as a second language. For these children school English is difficult to use since it conflicts with the dialect normally heard at home.

Subjects in Group Five are most unlike the students in Group Three in that the major differences for four of the language variables are significant (-19.0*, -18.7*, -15.9*, -9.1, -15.9*). While the difference in mean conventionality scores of -9.1 is not statistically significant, it almost exceeds the critical value of 9.48, suggesting that the profiles are essentially different on all five language variables. In many respects this is not surprising. The children in Group Five, all Negro, from culturally deprived environments, are notable as voluable, excited talkers. While the children of Group Three are non-talkers, those of Group Five could be easily classified as impulsive, motivated talkers.

Other differences among the groups stand out. As already indicated, children in Group Four are very similar to those in Group Two with respect to language style, but their deviations from the other groups do not parallel the differences for Group Two. For example, the differences in means for Group Two and Group Five are given by (-6.7, -14.5*, -8.3*, -10.1*, -6.2) while for the comparison of Group Four with Group Five, the vector of mean differences is given by (-7.8*, -9.5, -8.5*, -4.1, -6.4*). Group Two differs from Group Five with respect to conventionality while Group Four subjects are closer to Group Five on this dimension. The group most unlike Group Four is Group Eight. As shown in the fourth matrix of Table 17, the vector of mean differences for these two groups is given by (19.9*, 7.9, 11.9*, -12.6*, 16.2*). This great profile difference was also noted for the comparison of Group One and Group Two subjects with Group Eight. Since the subjects in Groups One, Two, and Four constitute the most proficient users of spoken language, the difference of Group Four from Group Eight is not unexpected.

Though Group Five subjects are all Negro, they nonetheless are quite different from Negro subjects included in Groups Six, Seven, and Eight. As has been pointed out, the similarity of their oral syntax to that of subjects in Group One is quite impressive; thus the differences in their profile from Groups Six, Seven, and Eight is not surprising.

Group Six students are most like Group Seven students. The vector of mean differences for these two groups is given by (-3.9, -12.0*, -2.7, -3.2, -2.5). This similarity is understandable when it is recalled that most of these subjects come from low SES homes, most are Negro, they hear either a foreign language or a social class dialect at home, and their IQ is below average. The major distinguishing feature between their profiles is the acute problem the students in Group Seven have with mazes, possibly resulting from poor self images and the tension between school language and home language.

As has already been suggested, Group Seven subjects could serve as the opposite to Group One subjects. They do show some similarities to the children in Group Eight with respect to dependent clauses and conventionality, but they are very much unlike them with respect to the remaining three language variables.

Summary Comments on the Similarities and Differences Between the Eight Language Groups.

In this section, a post hoc analysis was conducted on the differences between the profile means of the five language variables for the eight language groups generated on the early year's language data. The results of this Scheffé type of analysis substantiates the subjective analysis presented earlier in the narrative. In some respects this supporting evidence is not unexpected since the eight language groups were generated on the same data used for the post hoc investigation of the mean differences. According to this analysis the following conclusions were made.

Group One students (conventional English speakers who use the spoken word with power) stand out as the most effective users of conventional standard English. They are similar to Group Five students (minority students who use nonconventional English with power) with respect to fluency, dependent clauses, and elaboration index. In this sense, both groups use language with equal vigor and complexity. With respect to conventionality and freedom from mazes, Group One students are similar to Group Two students (conventional English speakers who use the spoken word in an unremarkable way), Group Three students (conventional English speakers who avoid verbal expression because of damaged self-images), and Group Four students (typical middle-class users of American English). Finally, Group One students are exceptionally unlike the Group Six students (students to whom standard English is a second dialect or foreign language), Group Seven students (students who are incomprehensible users of nonconventional English), and Group Eight students (unconventional English speakers who avoid the use of language).

While the language profile for the Group Two children is non-statistically different from the language profile of Group Four, qualitative differences exist between the groups with respect to intelligence and teacher ratings of oral language proficiency. While these children differ from all of the other groups on at least two of the five language characteristics, they appear to be most unlike the Group Eight students.

Group Three students are most like Group Six in that both groups tend to avoid verbal communication although for different reasons. Group Three subjects tend to be non-talkers because they tend to use English as a second language. Because they avoid spoken language, they are most unlike Group Five students who are notable talkers, and Group One students who demonstrate more fluency, dependent clauses, and elaboration of syntax.

Group Four subjects are very similar to Group Two subjects in their use of language but most unlike Group Eight subjects. Otherwise, this group appears to be intermediate to the remaining groups of the study.

Group Five children are quite different from Group Six, Seven, and Eight, even though all four groups are essentially Negro and from low socio-economic environments. The Group Five subjects exhibit better facility with communication than do the other three groups. Group Six children are very similar to Group Seven except that the Group Seven children have extreme problems with mazes resulting from their less than positive psychological security. Otherwise, Group Six students differ in large degree from the other groups.

Group Seven students are at a polar position to the Group One students.

Group Eight students are quite different from all the other remaining seven groups.

Language Profiles of the Average T-scores for the Tenth, Eleventh, and Twelfth Grades.

Over the years of schooling, the language groups become more like one another. In Table 18 the basic data and statistics for the eight language groups are summarized for the Tenth, Eleventh, and Twelfth Grades. If the statistics of

Table 18. Basic Statistics for the Standardized Language Variables of the Eight Language Groups at Grades Ten, Eleven, and Twelve.

Statistic	Variable	1	2	3	4	5	6	7	8	Average
Averages	Fluency	56.2	54.3	49.0	47.7	50.7	46.2	45.3	44.0	50
	Mazes	50.9	50.4	52.1	49.6	51.4	51.7	43.2	47.9	50
	Dep.Cl.	54.6	52.8	50.6	49.3	50.4	47.2	44.5	44.6	50
	Conven.	56.1	56.6	54.1	49.4	47.3	43.0	36.7	41.9	50
	Elab.Ind.	54.8	52.7	50.4	48.3	48.7	47.4	45.2	47.0	50
Standard Deviation	Fluency	7.7	8.4	7.6	7.6	6.1	8.2	4.9	8.4	7.8*
	Mazes	7.9	8.5	5.9	11.1	5.4	7.2	11.4	11.9	9.0*
	Dep. Cl.	7.6	6.8	5.5	7.4	6.4	7.9	5.9	7.2	7.0*
	Conven.	3.7	2.9	3.7	6.7	5.9	7.8	8.1	8.8	5.8*
	Elab. Ind.	8.1	7.8	5.7	7.6	4.7	7.1	5.3	8.0	7.2*
Sample Size		27	51	24	37	10	30	16	16	20.85**

*Within sample estimate of variance.

**Harmonic mean of the sample sizes.

Table 18 (Grades 10, 11, 12) are compared to the corresponding data of Table 12 (Grades 1, 2, 3), it is seen that the eight language groups have converged toward one another over the years, so that the large differences in profiles existing at the early grades are considerably reduced at the later grades. For example, at the first, second, and third grades the range in average fluency scores extended from 32.4 to 60.5 for an 18.1 range in standard units while the corresponding range for the tenth, eleventh, and twelfth grade data, the range in mean scores is from 44.0 to 56.2 for a 12.1 range in standard units--a reduction of 6 standard units. Corresponding reductions in maximum deviation between the group means are noted for the remaining four language variables.

Also, as shown by a comparison of the standard deviations of the two sets of data, the variation within the individual groups has increased with the exception of conventional usage, so that more overlap between groups is found at the advanced school years. For example, the range in standard deviations for the early grades fluency data was given by 3.6 to 6.6 while at the later grades, the corresponding range in standard deviations was given by 4.9 to 8.4. Thus, maturation, aging, formal education, social development, and time have acted to spread the members of a group from one another. With the spreading out of the subjects within a group and the general movement of the groups toward a common average position, it follows that the subjects across the total sample have become more homogeneous with respect to language.

By examining the language profiles of Figures 15 through 22, one can easily see this trend towards a more uniform language. If the profiles are examined group by group, it is seen that for Group One subjects, the peaks and troughs of the elementary school year profile have been reduced. However, the relative positions of the mean maze score and conventionality measure are identical to the early profile values: In essence, these two language characteristics have not been materially affected over time. Apparently, children who begin their life of oral communication with conventional speech patterns free from mazes have a head start on these two characteristics of spoken speech, an advantage they maintain as they age. In other words, it seems that acceptable or "good" speech habits developed at any early age are difficult to lose or damage as an individual ages.

For the Group Two subjects, the early and later profiles are very similar, but what is of greater interest is their similarity to the profile of the Group One subjects. While it was reported that Group One differed from Group Two with respect to color and power of language,it is seen that these differences have been eliminated by the time the high school years are reached. While the profiles are very close or colinear to one another, it is still true that the Group One profile is slightly above the Group Two profile. However, as will be seen, the variable by variable differences are not significant so for all practical purposes, the profiles are identical.

For the Group Three subjects, the peaks and troughs

Figure 15. Subjects with Broad-based Vocabularies, Superior, Fluent, Conventional English, and Effective Complex Syntactical Structures.

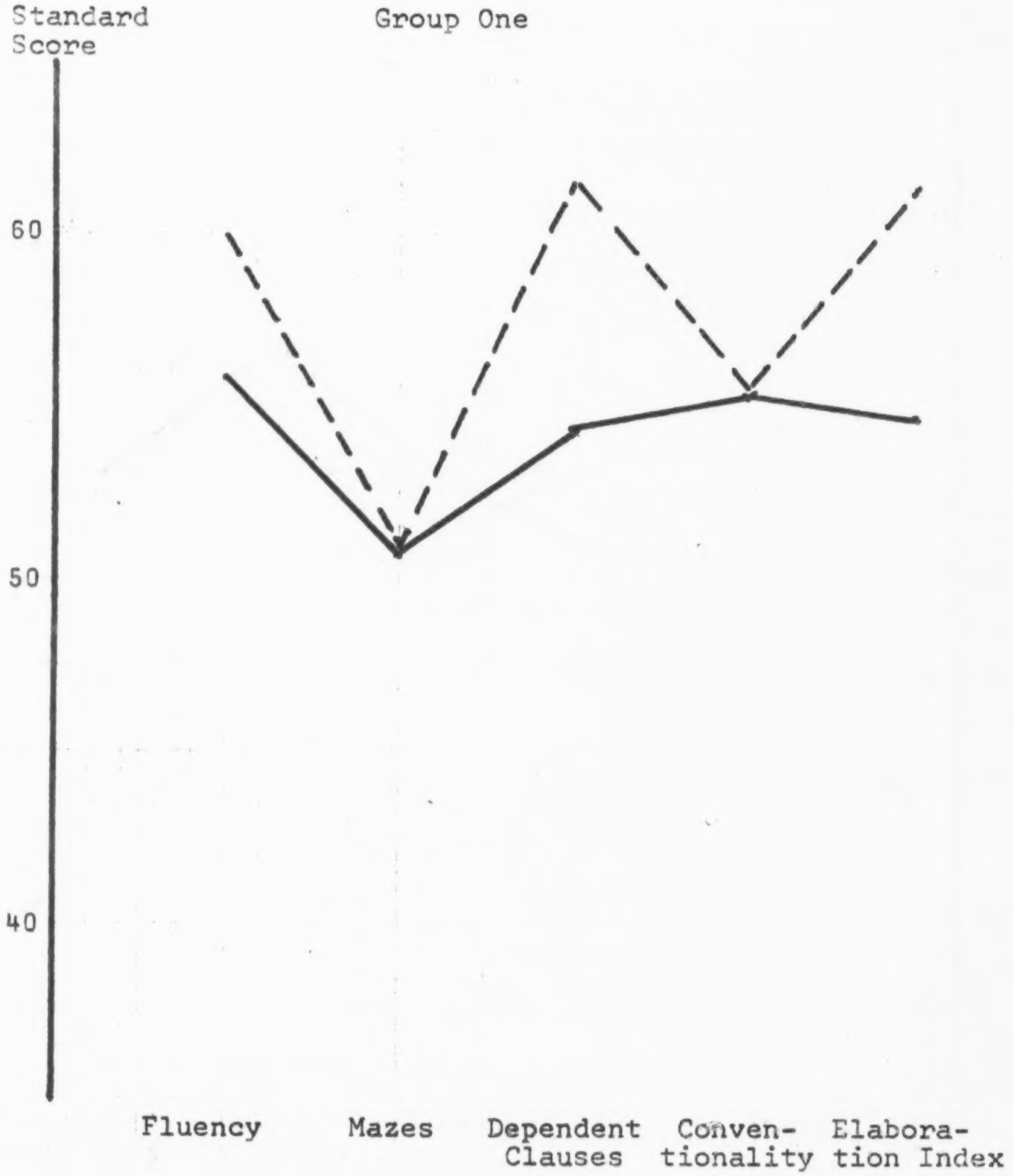


Figure 16. Students Who Use English in a Highly Conventional, Fluent, Coherent, but not Remarkably Superior Mode of Expression.

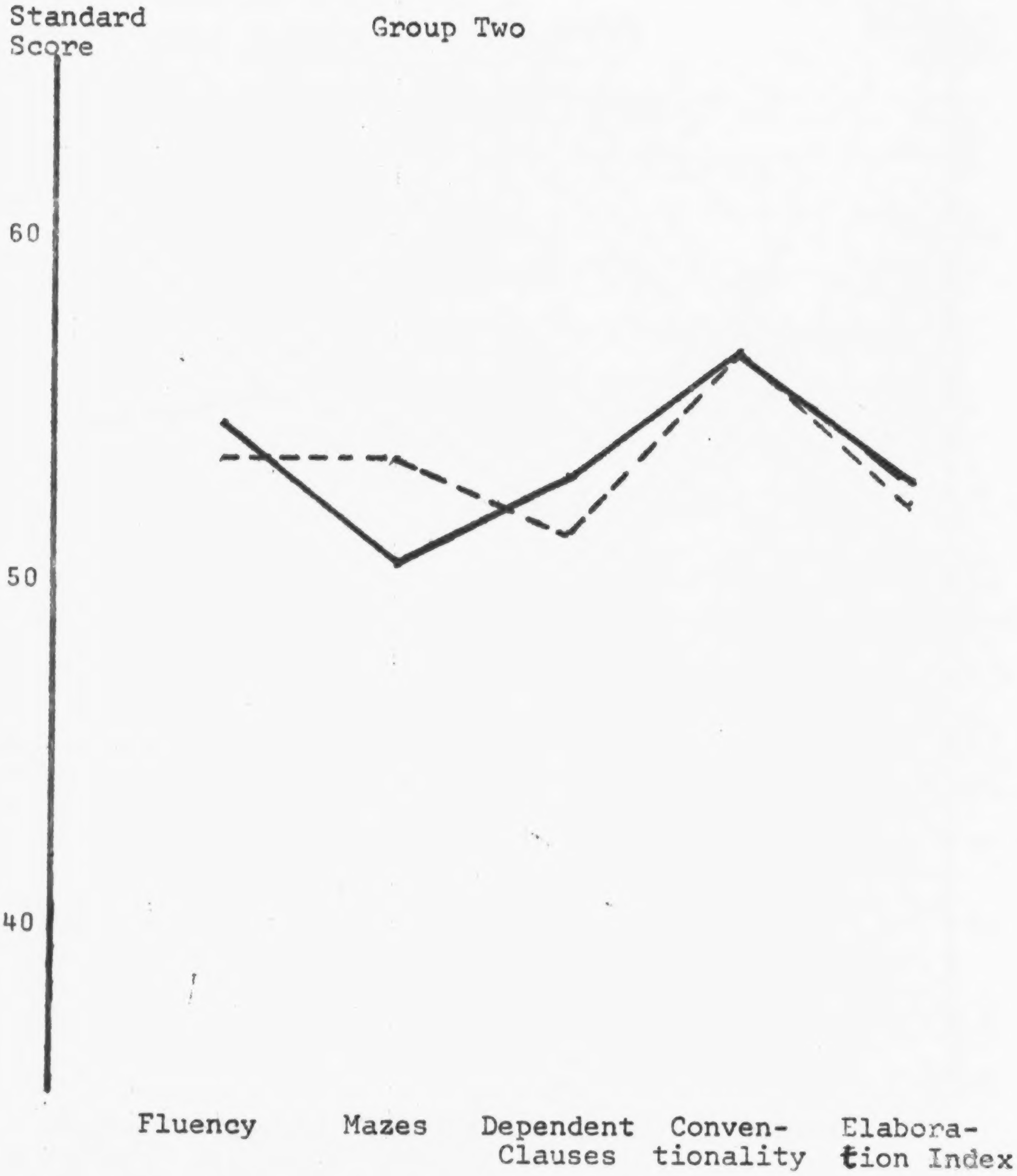


Figure 17. Students of Above Average Intellectual Ability Who, Because of Personality Shyness and Weak Self-images, Avoid Verbal Expression and Interchanges.

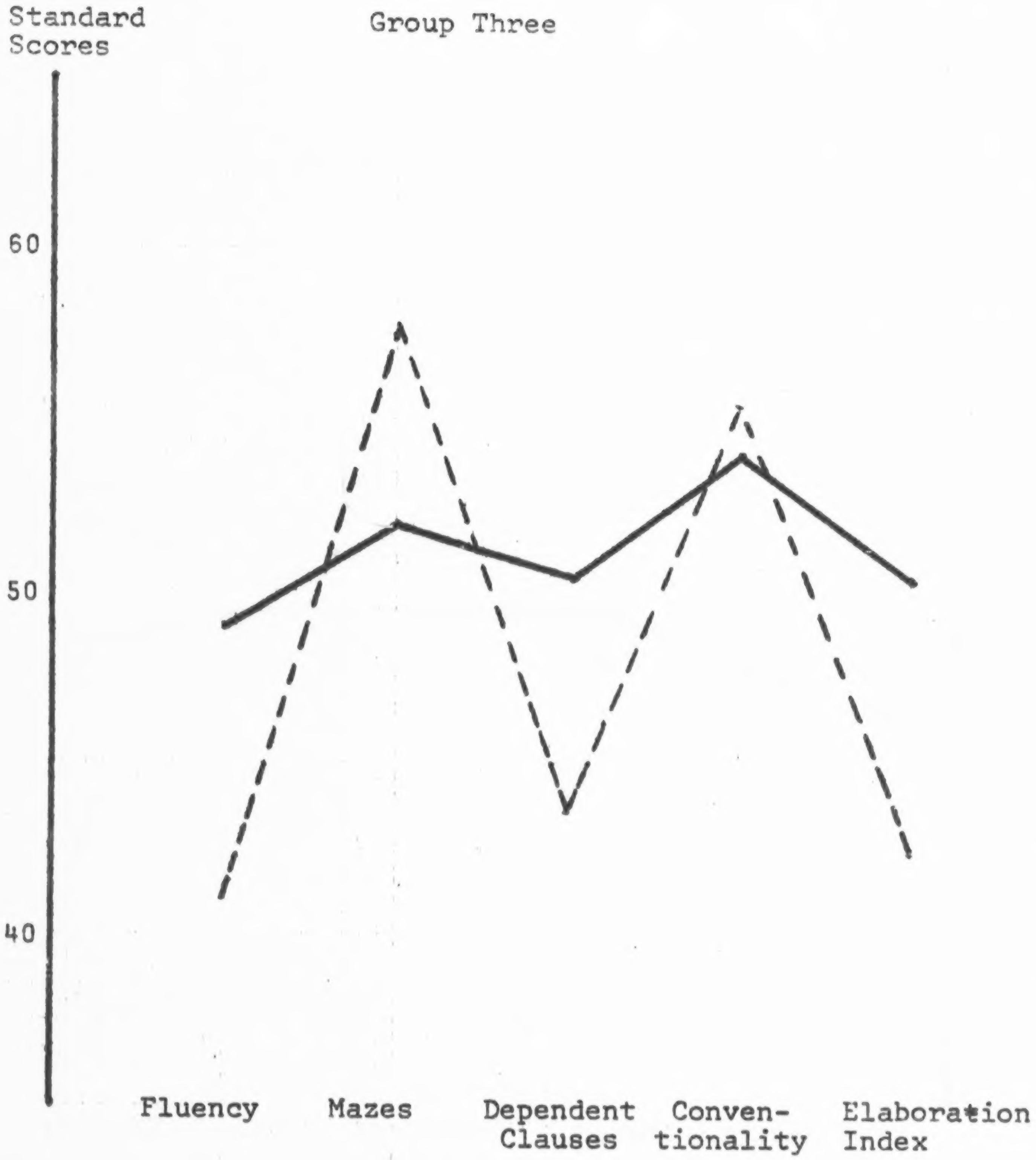


Figure 18. Typical Middle-class Users of American English.

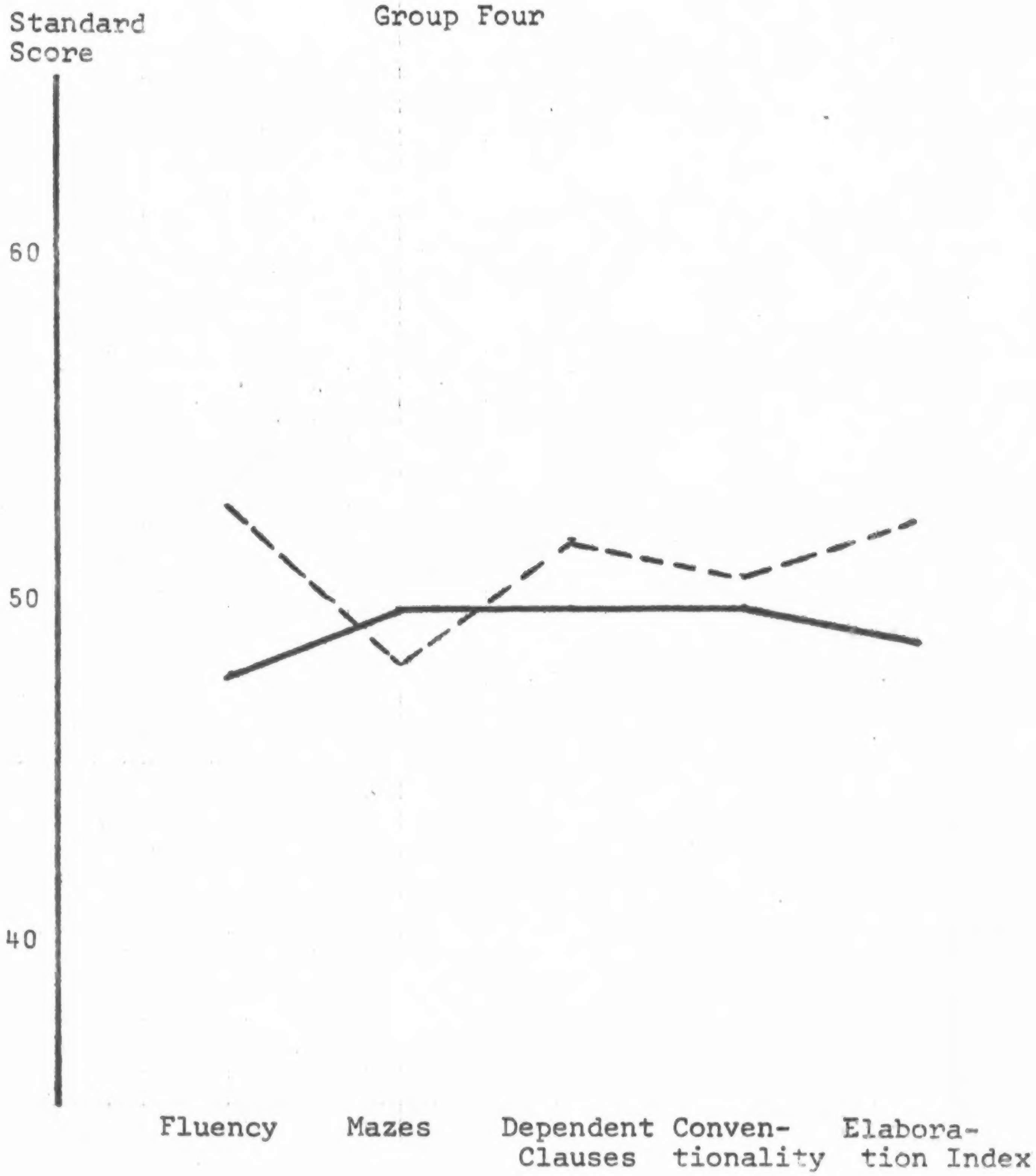


Figure 19. Talkative Subjects from Culturally Deprived Backgrounds Who Use Non-conventional Speech Patterns.

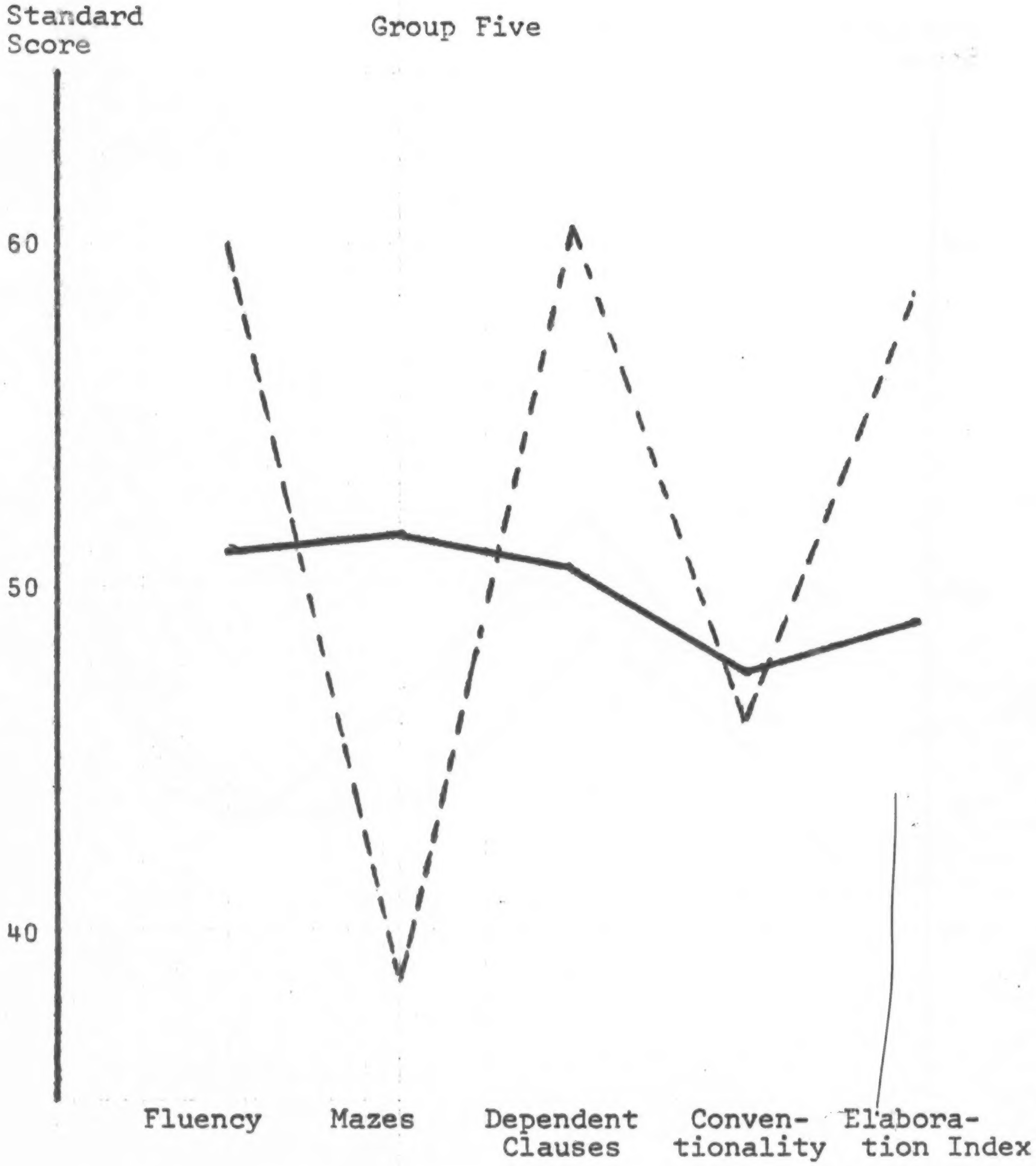


Figure 20. Children to Whom Standard English is a Second Dialect or a Foreign Language.

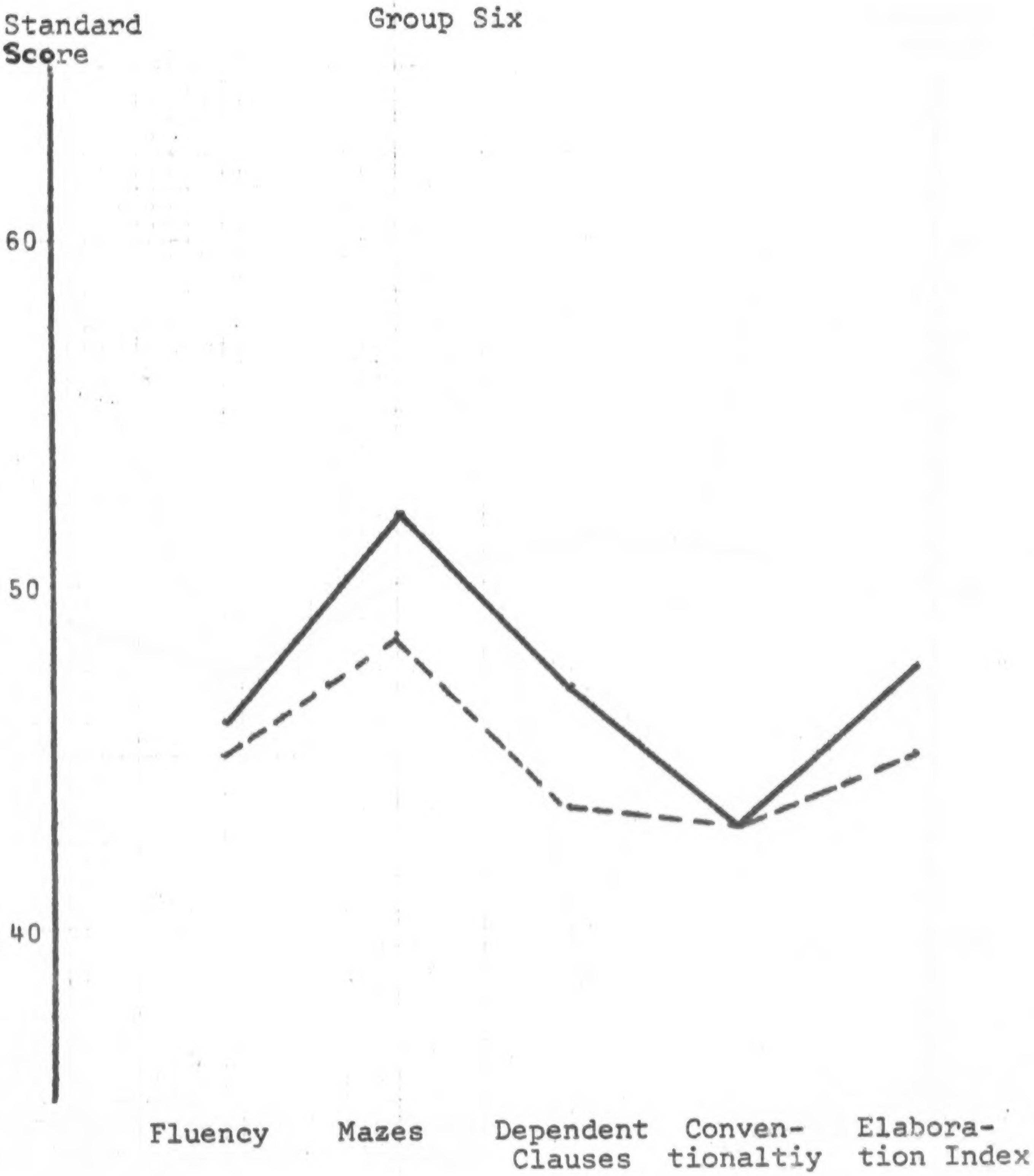


Figure 21. Pupils Whose Use of English is Non-Standard and Difficult to Comprehend--for Middle-class Teachers.

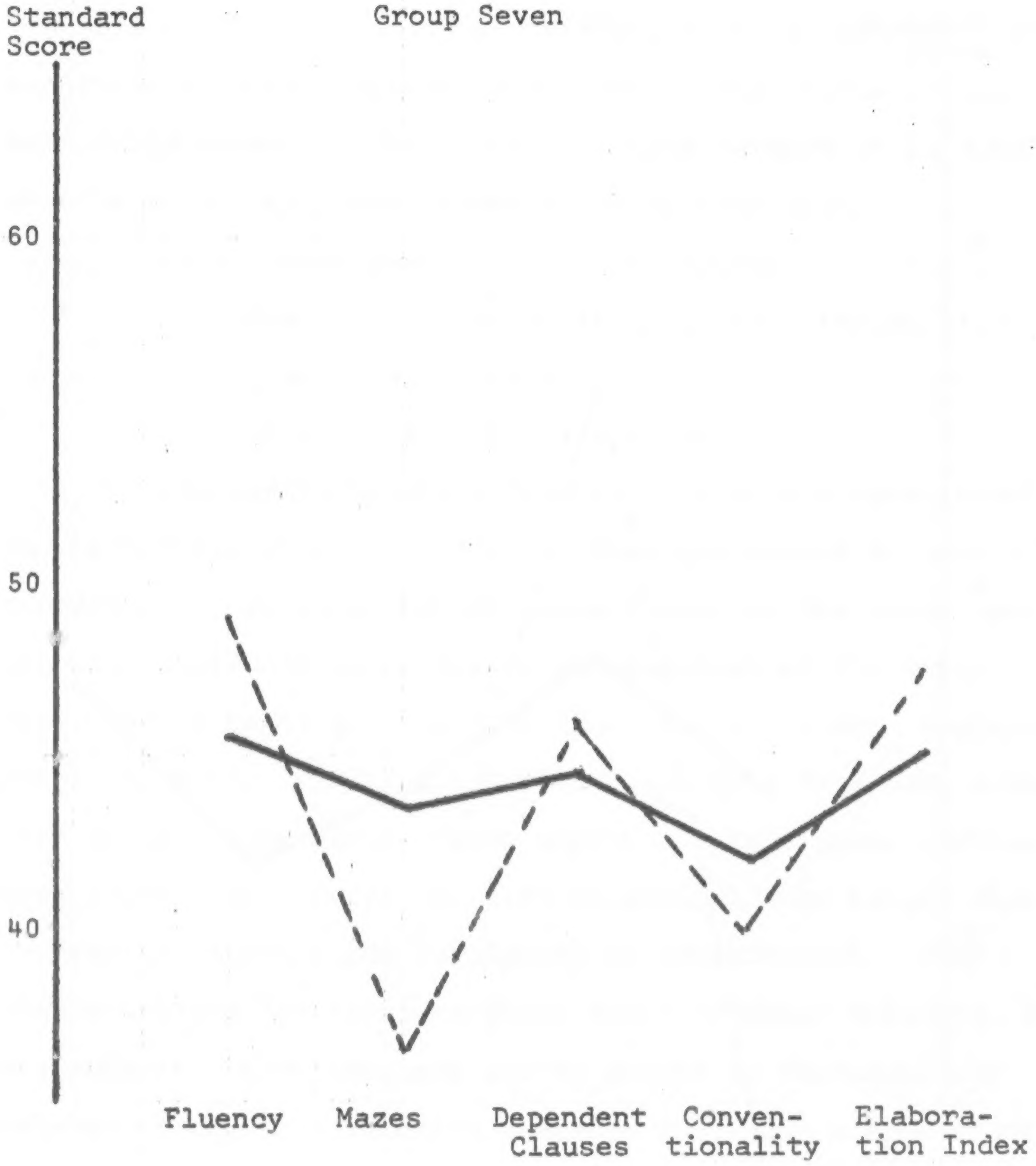
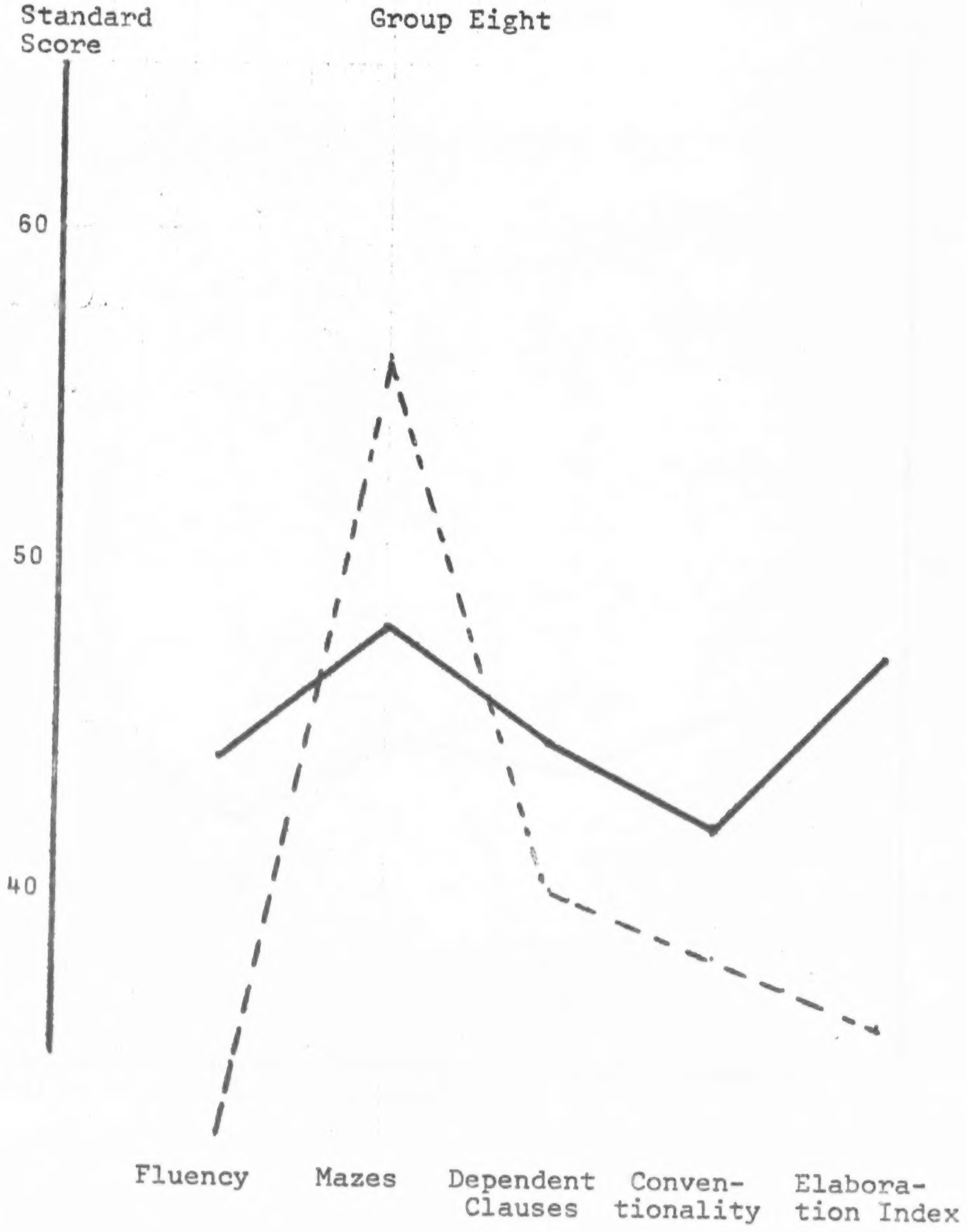


Figure 22. Students Who are Unconventional Non-language Users.



in the profile have also diminished and the once shy students comprising this group have progressed to a language usage position that is more in line to that of the students in Group One and Group Two. Without doubt, the large changes in language proficiency made by these students is striking. The magnitude of these changes is evident in the vector of the mean differences for the early and later periods of language experiences. For these students, it is seen that:

$$\begin{aligned} \Delta &= (\text{later period}) - (\text{early period}) \\ &= (49.0,\ 52.1,\ 50.6,\ 54.1,\ 50.4) - (41.2,\ 57.3, \\ &\qquad 44.0,\ 55.4,\ 42.6) \\ &= (6.8,\ -5.2,\ 6.6,\ -1.3,\ 7.8) \end{aligned}$$

According to these figures, the only characteristic on which these students failed to show any change is conventionality. They were already conventional at the early ages and they continued to be highly conventional at the later ages. As is recalled, the high IQ of these children corresponded to their initial conventional and free from maze language style. Apparently, their superior intelligence combined with school experiences and life in general have helped them improve or increase the complexity of their speech. When characterizing the early language style of these children, it was suggested that language growth should be exponentially related to use. It is quite possible that such a growth was experienced by these students. As these high IQ students aged, one would expect them to increase the complexity of their expression and expand it on a superstructure consisting of a speech pattern that is already free of mazes and highly conventional in form. If this were to have occurred, it would have to be concluded that for these kinds of children, formal English instruction in school will be effective and rewarding, especially if it is well done.

The subjects in Group Four and Five have become more alike in language. Those in Group Four continue to have in the secondary school a profile paralleling the 50th percentile value across all five language variables. As was noted earlier, these students defined the median language group at the early ages, and as is now apparent, these same students have the continued distinction of defining the median language group at the later ages.

The later profile for the Group Five students is considerably unlike the profile noted at the earlier ages in that the extreme peaks and troughs have been thoroughly plained. The mean differences in profiles for this group are given by:

$$\begin{aligned}\Delta &= (\text{later period}) - (\text{early period})\\ &= (50.7,\ 51.4,\ 50.4,\ 47.3,\ 48.7) - (60.2,\ 38.6,\\ &\qquad 60.0,\ 42.2,\ 58.5)\\ &= (-9.5,\ 12.8,\ -9.6,\ 5.1,\ -9.8)\end{aligned}$$

As these figures suggest, this group made the greatest changes over the ten to twelve year period of the study. Their fluency, their use of dependent clauses, and their tendency to elaborate, relative to the other students in the study, decreased by about one standard deviation. Where they were once among the most fluent, complex speaking students, they are at the later period, near the median. This does not mean they have undergone a deterioration in complexity of language because it must be recalled that the early profile is defined in terms of the language among all the children at the early ages while at the later ages, the language is defined in terms of the language at the time of the data collection. It should be emphasized that this study is based upon relative and not absolute comparisons since all data has been converted to T-scores. In any case, it is also apparent that these talkative Negro subjects have brought their maze problems under control and have adopted a more standard American speech pattern. Finally, it should be noted that the profile for the Group Five students is very similar to the Group Four subjects who at the early ages represented typical middle-class users of American English.

For the Group Six subjects, the similarity of their later age profile to the early age profile is striking. It suggests that aging, schooling, maturation, and experience has not altered their language to any appreciable degree.

The later profile for Group Seven seems to have moved to a lower level on all variables except mazes. While it appears that their problems with mazes has improved, their use of language is still quite poor.

Finally, the Group Eight subjects seem to have improved considerably since their later profile is similar to the profile of the students in Group Six. It should be remembered that this group contained two groups of pupils whose language was alike for dissimilar reasons. One sub-group came from homes using a foreign language and the other sub-group was limited in ability.

Examination of the standard deviations reported in Table 18 with the corresponding statistics reported in Table 12 suggests that the students within individual language groups have become more variable with respect to language. In addition, mazes tend to show the greatest variation since the within sample standard deviation of nine units nearly equals the unconditional standard deviation of 10 units. For Groups Four, Seven, and Eight, the standard deviation for mazes exceeds the value of the basic T-variables. However, Groups One, Two, and Three tend to remain homogeneous with respect to conventionality. These values are just slightly larger than the values reported at the early ages.

Summary Comments on the Later Age Profiles of the Eight Language Groups

Over the years spanned by this study, many of the initial language differences that existed between the groups at the early ages have been diminished in magnitude, and in many cases, have been completely eliminated. Furthermore, the students within the individual language groups have fanned out from one another so that the boundaries between the language groups overlap one another to a considerable degree.

Students in Group One maintained their advantage of speaking in a style free of mazes and conventional in form. While they are still successful with fluency, dependent clauses, and elaboration, their position relative to the remaining students is not as extreme as it was at the early ages.

Students in Group Two have progressed to a point where their language profile is almost identical to that of the Group One students.

Group Three students have experienced remarkable changes in language profiles to one characterized by greater fluency, use of dependent clauses, and a tendency to elaboration. While these students were very shy nontalkers at the early ages, they developed a positive style on a firm foundation of conventional English free of mazes.

Group Four students maintained their relative standing as the median group. This does not mean they failed to grow in use of language. Since all comparisons in this study are relative to the total sample average, it follows that this group continues to define the norm even though their oral speech is more mature than it was in the early years of elementary school training.

The ten Negro students of Group Five made the greatest changes in oral language over the years covered in this study. Their problems with mazes were dramatically reduced and their movement to a more conventional mode of speech is quite apparent. While their relative superiority with respect to fluency, dependent clauses, and elaboration has disappeared--like that of Group One--their use of these elements of spoken English is still quite effective. Because of these changes in spoken language over the years, their later profile has come to resemble that of the Group Four students who basically represent average middle-class users of American English.

For Group Six students, the similarity of their early and later age profiles is striking. This suggests that aging, schooling, maturation, and experience has not altered their language style to any appreciable degree.

The relative position of the Group Seven students appears to have deteriorated over time with respect to all language variables except freedom from mazes. Their original use of oral language was poor and at the later ages is even poorer when compared to that of the other students.

Group Eight students have shown some growth in language proficiency in that their oral language use is similar to that of Group Six students.

Multivariate Analysis of the Average T-scores for the Tenth, Eleventh, and Twelfth Grade Data

The graphic representations of the distributions of the five language variables for the eight language groups at grades ten, eleven, and twelve are shown in Figures 23 to 27. When these graphs are compared to the graphs of Figures 9 through 13, it is seen that the large variation existing between the groups at the first, second, and third grades has been virtually eliminated on a number of the language characteristics. This movement to a common mode of oral and verbal expression is examined in the remaining narrative of this section.

The correlation matrix for the average standardized language variables for grades ten, eleven, and twelve is as shown in Table 19. As with the correlations found for the early years' groupings, mazes and conventionality tend to be non-correlated with any of the remaining language variables. This suggests that the information collected on these two variables at these advanced age levels is providing unique

Figure 23. Histograms of the Fluency Measures for the Eight Language Groups at the Later Ages.

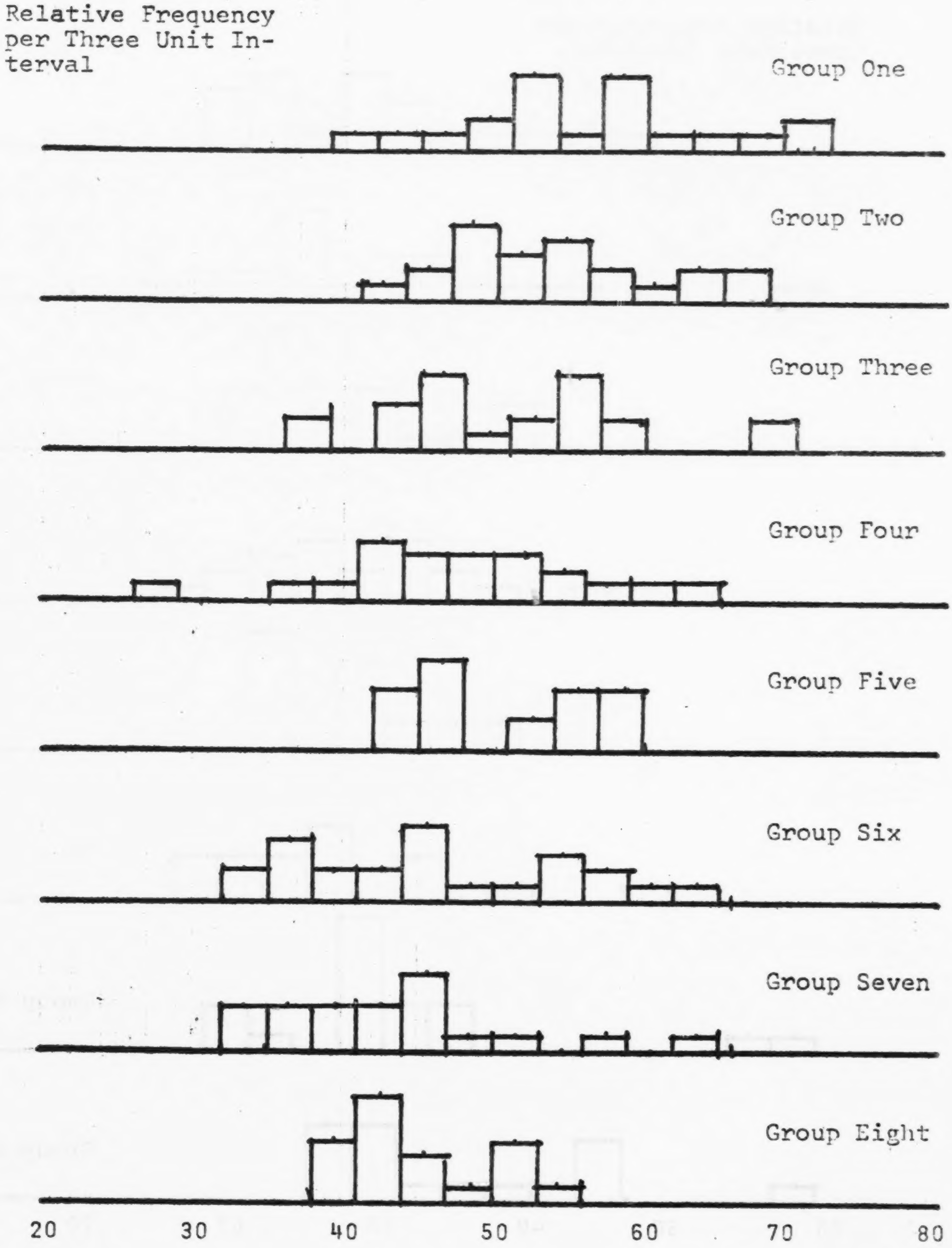


Figure 24. Histograms of the Freedom from Mazes Measures for the Eight Language Groups at the Later Ages.

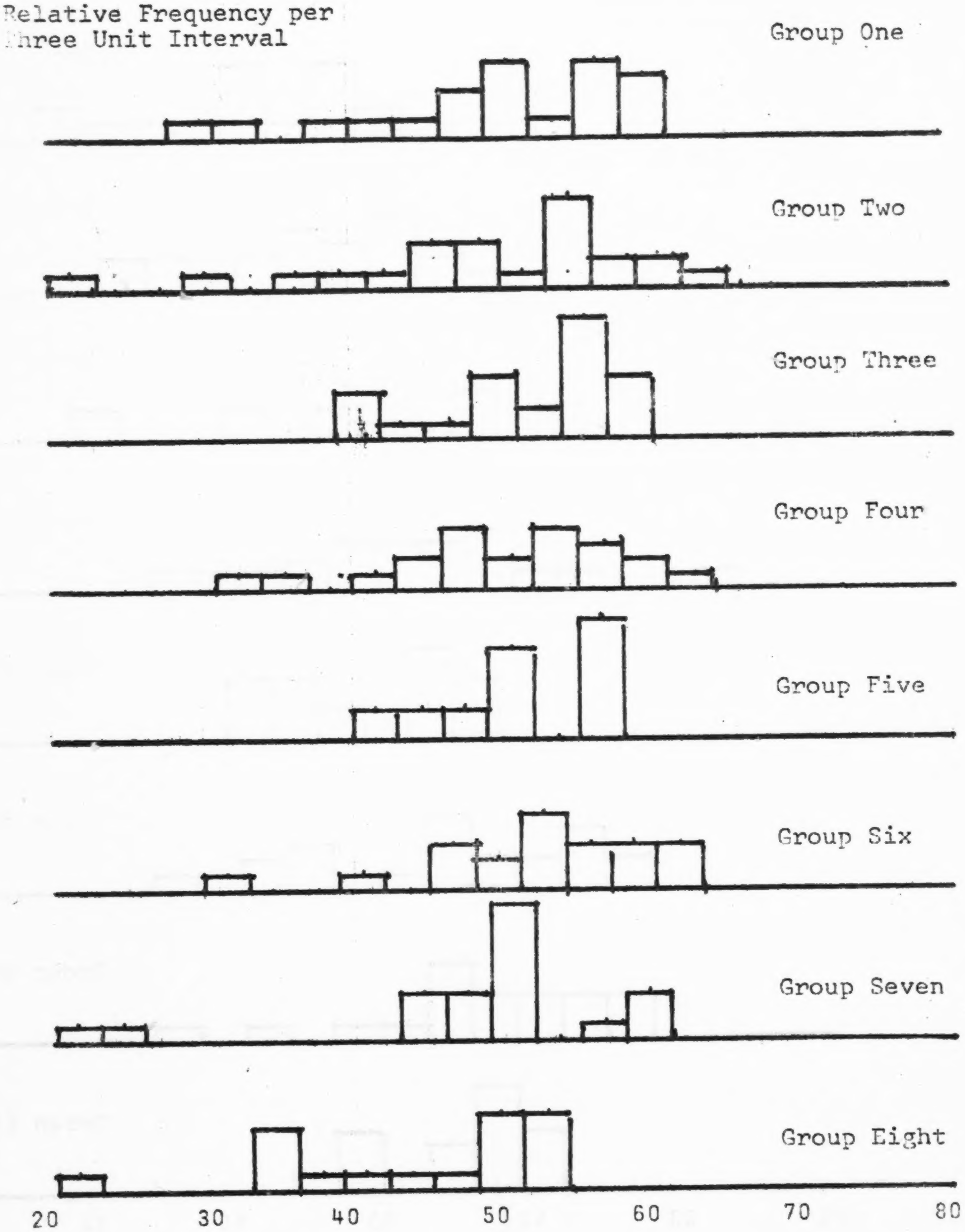


Figure 25. Histograms of the Dependent Clause Measurement for the Eight Language Groups at the Later Ages.

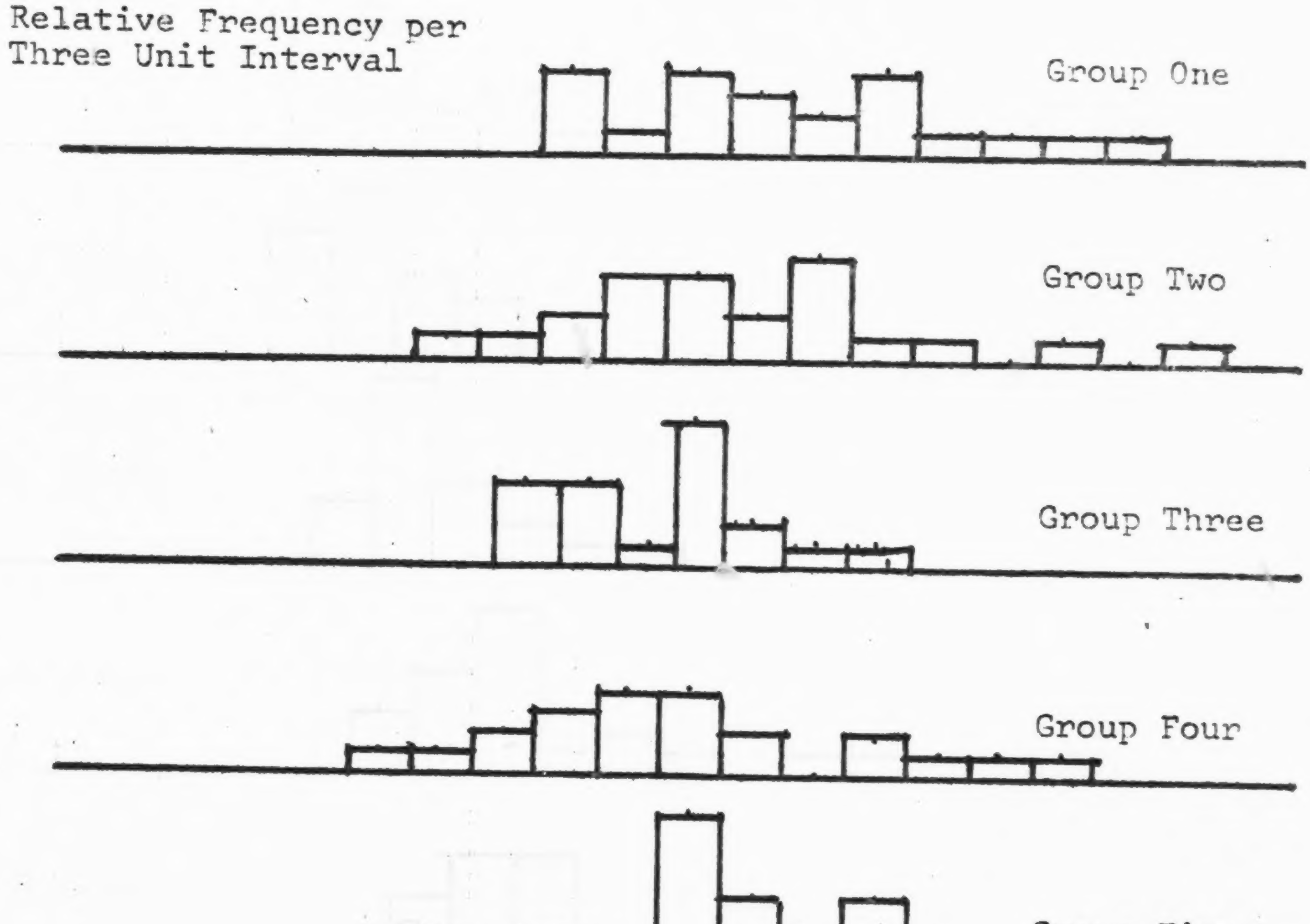


Figure 26. Histograms of the Conventionality Measures for the Eight Language Groups at the Later Ages.

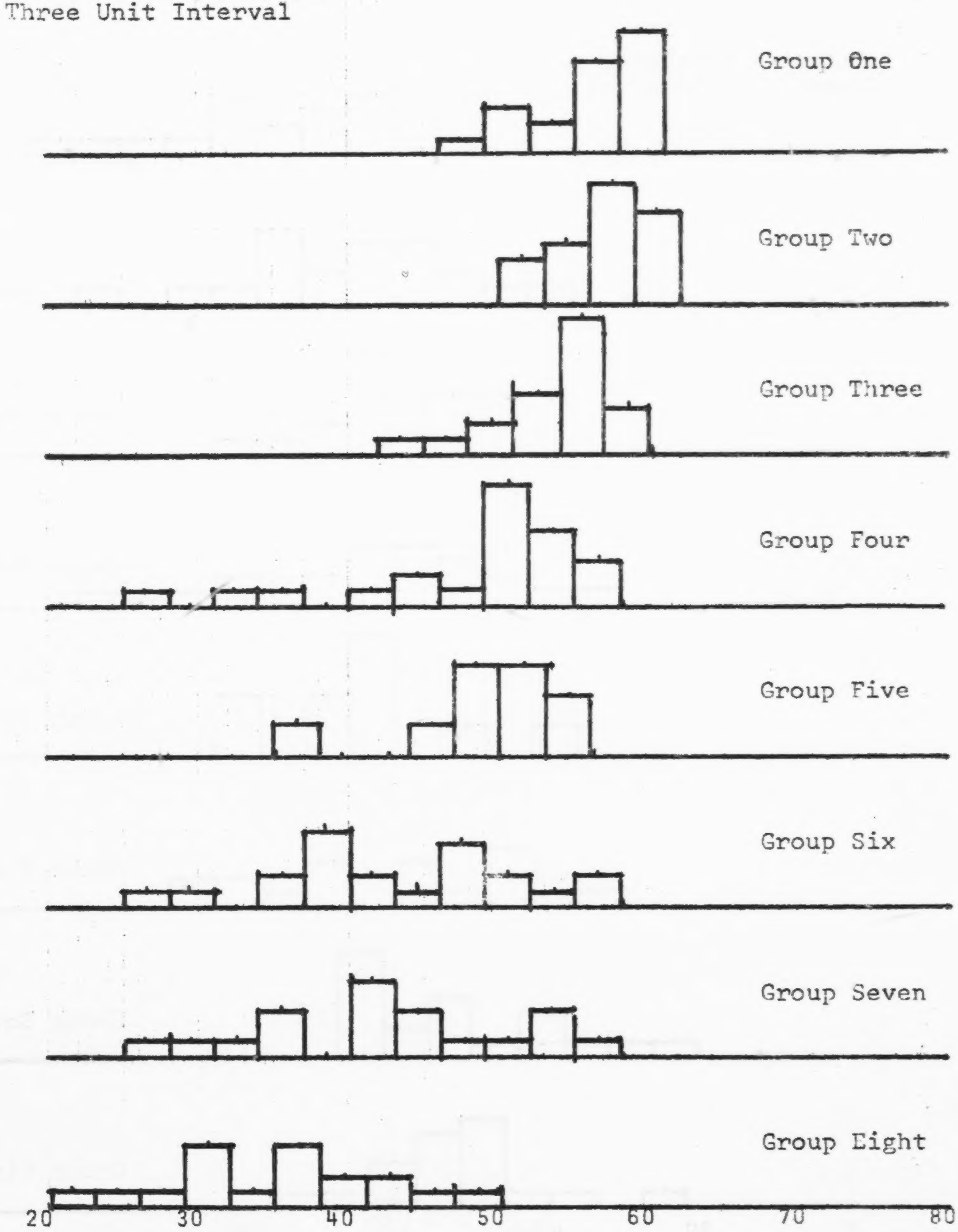


Figure 27. Histograms of the Elaboration Index for the Eight Language Groups at the Later Ages.

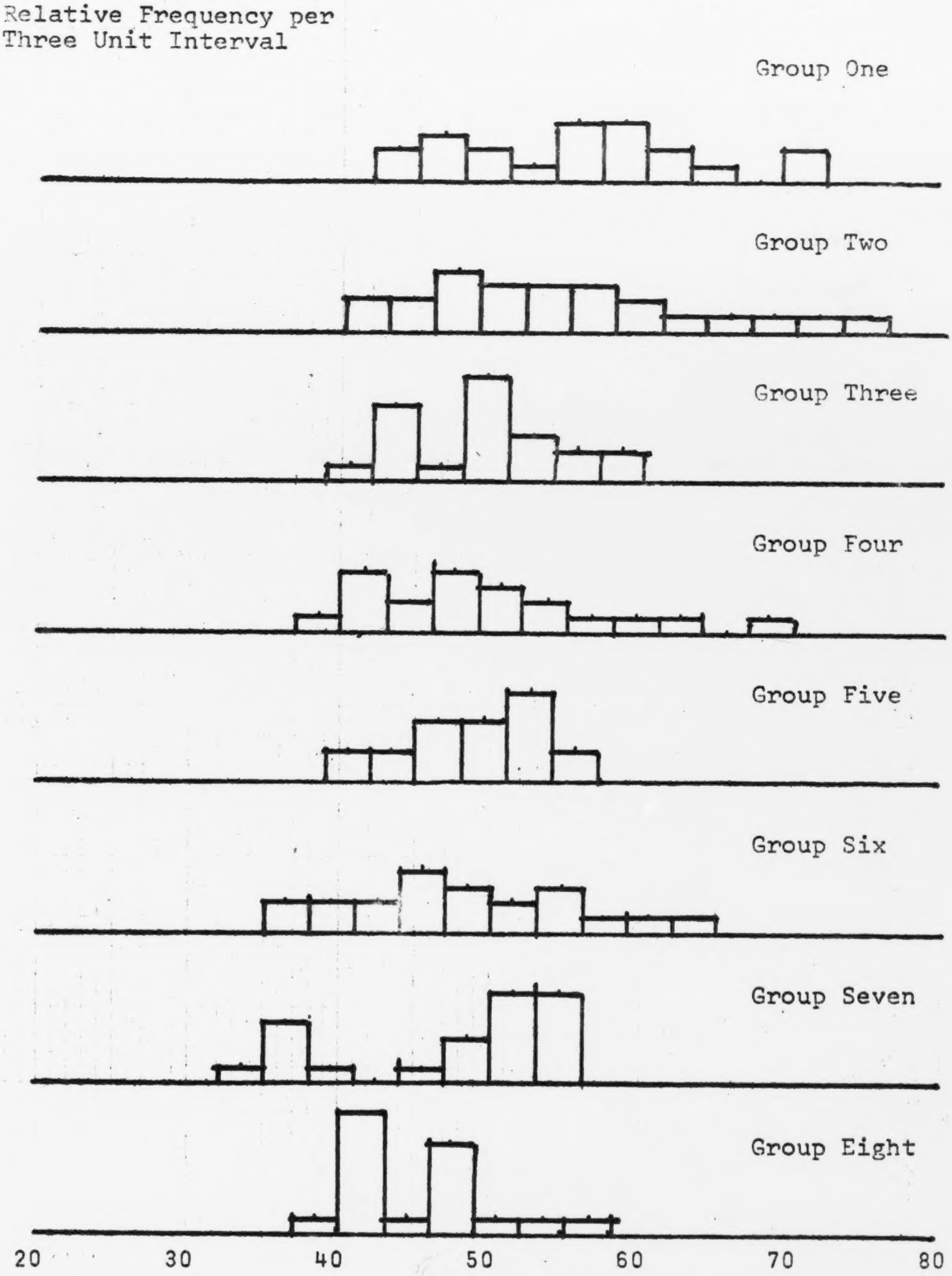


Table 19. The Within Correlation Matrix for the Standardized Language Variables at Grades Ten, Eleven, and Twelve.

Variable	Fluency	Mazes	Dependent Clauses	Conventionality	Elaboration Index
Fluency	1.00	-.02	.68	.14	.75
Mazes		1.00	.10	.03	.08
Dependent Clauses			1.00	.10	.83
Conventionality				1.00	.09
Elaboration Index					1.00

measures of language characteristics.

However, the high within group correlations between fluency, dependent clauses, and elaboration index suggests that a considerable amount of redundancy is being observed on these three language characteristics since the correlations between fluency and dependent clauses is equal to .68, between fluency and elaboration index is .75, and between dependent clauses and elaboration index is .83. These three correlations are considerably higher than the unconditional correlations reported in Table 1 where the average corresponding correlations were equal to .43, .57, and .53, suggesting that at older ages these variables are more related to some common element of language that is different from that encountered at the early ages.

On the other hand, it also follows that the differences between the groups have been reduced since the within correlations of .68, .75, and .83 are almost identical to the total correlations of .74, .78, and .85 reported in Table 4. Thus, the within group correlations are essentially identical to the total correlations.

In this sense, the within correlation analysis agrees with the canonical analysis reported for the total correlations. The conclusions drawn from that analysis for the total group of students can be extended to each of the individual eight groups of this study. Furthermore, the similarity of Table 19 to the lower portion of Table 4 suggests that group differences in fluency, mazes, dependent clauses, and elaboration index are essentially zero. As will be seen, this is the case.

The B and W matrix for the observed data at the later grades is shown in Table 20. With these matrices, the value of the F-ratio, as approximated by the Box procedure, is given by $F = 6.48$. This F-value, considerably less than the F-ratio for the first, second, and third grade statistics ($F = 26.56$), suggests that the differences between the later age profiles are considerably less than those observed at the earlier ages. As with the early age analysis, $\nu_1 = (G - 1)p = (8 - 1)(5) = 35$, and $\nu_2 = 839 \sim \infty$, so that the same decision rule is appropriate for this later age analysis. With $\alpha = .05$, the hypothesis of identical group profiles is rejected since $F = 6.84 > 1.43$. Thus, it is concluded that the language profiles are statistically different from one another on at least one of the language variables.

For these data, the numerical value of the Roy criteria is given by $\theta = .5926$. Since this number exceeds the $\alpha = .05$ critical value of $x_\alpha = .11$, the hypothesis of identical language profiles is rejected. This agrees with the decision made on the basis of the Box form of the F-test. The univariate analysis of variances for the five language variables are shown in Table 21. Any F-ratio exceeding 2.08 is indicative of a statistically significant difference among the means of the corresponding variables.

Now that the hypothesis of identical profiles has been rejected, the next task is to identify the significant differences among the means of the eight language groups for the five language variables. As can be seen, the differences

Table 20. The B and W Matrices for the Tenth, Eleventh, and Twelfth Grade Average T-scores.

Matrix	Variable	Fluency	Mazes	Dependent Clauses	Conventionality	Elaboration Index
B	Fluency	509.4	-103.8	382.0	-714.6	347.6
	Mazes		152.2	-119.8	256.9	-95.7
	Dependent Clauses			310.2	-602.3	271.1
	Conventionality				1298.4	-533.7
	Elaboration Index					256.1
W	Fluency	12234.8	337.0	7525.2	-1258.6	8534.1
	Mazes		16321.2	-1337.8	278.1	1126.7
	Dependent Clauses			9955.1	-868.8	8576.8
	Conventionality				6934.5	690.2
	Elaboration Index					10645.3

Table 21. Univariate Analysis of Variance for the Five Variables at Grades Ten, Eleven, and Twelve.

Variable	Source of Variance	d/f	Mean Square	F-ratio	$\hat{\omega}^2$	Decision
Fluency	Between Groups	7	509.4	8.45	.23	Reject
	Within Groups	203	60.3			
	Total	210				
Mazes	Between Groups	7	152.2	1.89	.06	Do not reject
	Within Groups	203	80.4			
	Total	210				
Dependent Clauses	Between Groups	7	310.2	6.33	.18	Reject
	Within Groups	203	49.0			
	Total	210				
Conventionality	Between Groups	7	1298.4	37.96	.57	Reject
	Within Groups	203	34.2			
	Total	210				
Elaboration Index	Between Groups	7	256.1	4.88	.14	Reject
	Within Groups	203	52.4			
	Total	210				

among the means for the mazes are strictly due to chance factors, ($F = 189 < 2.08$). This is also suggested by the near zero value of the measure of explained variance for this variable which is given by $\hat{\omega}^2_M = .06$. The differences that existed among the groups with respect to problems with mazes at the earlier school years are no longer a significant source of variance existing among the groups. On this particular variable, the eight language groups have definitely come together. It appears that as students age, the problems they have with mazes become similar across the groups. By the time the late teens are reached, mazes are not a language variable differentiating among the groups.

As suggested by the remaining F-ratios and the $\hat{\omega}^2$ measures of Table 21, most of the differences should be associated with the conventionality measures since the amount of explained variance between the groups on this variable is given by $\hat{\omega}^2_C = 57\%$. Since $\hat{\omega}^2_F = 23\%$, $\hat{\omega}^2_D = 18\%$, and $\hat{\omega}^2_E = 14\%$, some differences, but not many, should be identified with these variables.

For this post hoc analysis, S continues to equal 5.01. On the basis of this numerical value, the corresponding critical values for the pairwise differences shown in Table 22 under the column heading C_p were determined. If these critical values are compared to the corresponding figures shown in Table 16, it is seen that before a significant source of variance can be identified, the difference between any two group averages must be considerably larger than the values determined for the first, second, and third grade statistics. However,

Table 22. Values of C , the Critical Value for the Pairwise Differences in Means for the Five Basic Language Variables at Grades Ten, Eleven, and Twelve.

Variable	MSW_p	$\frac{2}{N}\,MSW_p$	C_p
Fluency	60.2723	5.7807	12.05
Mazes	80.3966	7.7108	13.91
Dependent Clauses	49.0423	4.7037	10.86
Conventionality	34.1608	3.2764	9.07
Elaboration Index	52.4367	5.0292	11.24

this does not hold for mean differences with respect to conventionality. In any case, fewer significant findings are going to be found for the individual pairwise differences. This reduction in the absolute number of significant differences corresponds to the smaller F-ratio and smaller $\hat{\omega}^2$ values observed for the tenth, eleventh, and twelfth grade data.

Scheffé Type Analysis of the Differences Between the Mean Vectors of the Eight Language Groups at Grades Ten, Eleven, and Twelve

It appears that over a period of time many of the mean differences have been neutralized and pretty well eliminated as language discriminators. The pairwise differences in mean values for the five dependent variables across the eight language groups are presented in Table 23. In the first matrix of the table of mean differences, Group One is compared to each of the remaining seven language groups. As is immediately apparent, the large mean differences that existed between Group One students and the remaining seven language groups at the primary grades are not evident for these students at the high school grades, except for the conventionality measure.

However, the language variable we call fluency--length of communication unit--deserves special attention. Group One students still continue to differ from Group Eight students with respect to fluency. However, although the differences between the Group One and Group Six and Seven students on this variable exceed 10 standardized units (or one standard deviation), they are not large enough to be reported

Table 23. Variable by Variable Pairwise Comparisons Between the Profiles of the Eight Language Groups.

	1	2	3	4	5	6	7	8
Group One Versus the Seven Remaining Groups								
Fluency		1.8	7.1	8.4	5.5	10.0	10.9	12.1*
Mazes		-.5	1.3	-1.2	.5	.9	-7.6	-3.0
Dependent Clauses		2.1	4.3	6.4	6.1	7.3	9.5	10.0
Conventionality		.5	-2.0	-6.7	-8.8	-13.1*	-19.4*	-14.2*
Elaboration Index		1.8	4.0	5.2	4.2	7.3	10.0	7.7
Group Two Versus the Seven Remaining Groups								
Fluency			5.3	6.6	3.7	8.2	9.0	10.1
Mazes			1.8	-.8	1.0	1.4	-8.2	-2.5
Dependent Clauses			2.2	4.3	4.0	5.2	7.4	8.2
Conventionality			-2.6	-7.2	-9.3*	-13.7*	-19.9*	-14.7*
Elaboration Index			2.2	3.5	2.4	5.6	8.3	5.6
Group Three Versus the Seven Remaining Groups								
Fluency				1.3	-1.6	2.9	3.7	5.0
Mazes				-2.6	-.8	-.4	-9.0	-4.3
Dependent Clauses				2.0	1.8	3.0	5.2	6.0
Conventionality				-4.7	-6.8	-11.1*	-17.4*	-12.2*
Elaboration Index				1.2	.2	3.3	6.1	3.4

Table 23. (Continued)

	1	2	3	4	5	6	7	8
Group Four Versus the Seven Remaining Groups								
Fluency					-2.9	1.6	2.4	3.7
Mazes					1.7	2.2	-6.4	-1.7
Dependent Clauses					-.3	1.0	3.2	4.7
Conventionality					-2.1	-6.4	-12.7*	-7.5
Elaboration Index					-1.1	2.1	4.9	1.3
Group Five Versus the Seven Remaining Groups								
Fluency						4.5	5.4	6.6
Mazes						.4	-8.1	-3.5
Dependent Clauses						1.3	3.5	5.8
Conventionality						-4.4	-10.6*	-5.4
Elaboration Index						3.2	5.9	1.6
Group Six Versus the Seven Remaining Groups								
Fluency							.9	2.2
Mazes							-8.6	-3.9
Dependent Clauses							2.2	2.6
Conventionality							-6.2	-1.1
Elaboration Index							2.8	.4
Group Seven Versus the Seven Remaining Groups								
Fluency								1.3
Mazes								4.7
Dependent Clauses								-.1
Conventionality								5.2
Elaboration Index								-1.8

as statistically significant. Because of the numerical magnitude and from a practical and subjective point of view it is difficult to accept this lack of significance as a non-reliable difference. To test the possibility that fluency does differentiate between the groups so that the observed difference is indeed reliable, a complex contrast $\hat{\Psi}$, involving the appropriate sample means can be computed and then tested for significance by comparing it to S = 5.01. If $t = \frac{|\hat{\Psi}|}{SE_{\Phi}} > 5.01$, it is concluded that the tested difference is significant. For the observed data, the contrast for Group One versus the combined Groups Six, Seven, and Eight is given by:

$$\begin{aligned}\hat{\Psi}_F &= \bar{x}_{1F} - \left[\frac{30\bar{x}_{6F} + 16\bar{x}_{7F} + 16\bar{x}_{8F}}{62}\right]\\ &= 56.2 - \left[\frac{30(46.2) + 16(45.3) + 16(44.0)}{62}\right]\\ &= 56.2 - 45.4\\ &= 10.8\end{aligned}$$

The squared standard error of $\hat{\Psi}_F$ is given by:

$$SE^2_{\hat{\Psi}_F} = MSW_F\left[\frac{1}{27} + \frac{1}{62}\right] = (7.8)^2\left[\frac{1}{27} + \frac{1}{62}\right] = 3.2336$$

so that:

$$t = \frac{10.8}{\sqrt{3.2336}} = 6.00$$

Since t = 6.00 > 5.01, it is concluded that with respect to fluency, Group One differs from Groups Six, Seven, and Eight treated as a single collected group.

Corresponding to the fluency measure, it appears that Group One might differ from the combined Groups Six, Seven, and Eight with respect to dependent clauses and elaboration index. For dependent clauses, the associated contrast is given by:

$$\hat{\Psi}_D = 54.6 - \left[\frac{30(47.2) + 16(44.5) + 16(44.6)}{62}\right]$$

$$= 54.6 - 45.8$$

$$= 8.8$$

The squared standard error of $\hat{\Psi}_D$ is given by:

$$SE^2_{\hat{\Psi}_D} = MSW_D\left[\frac{1}{27} + \frac{1}{62}\right] = (7.0)^2\left[\frac{1}{27} + \frac{1}{62}\right] = 2.6044$$

so that:

$$t = \frac{8.8}{\sqrt{2.6044}} = 5.43$$

Since $t = 5.43 > 5.01$, it is concluded that with respect to dependent clauses, Group One differs from Group Six, Seven, and Eight treated as a single collective group.

For elaboration index the contrast is defined by:

$$\hat{\Psi}_E = 54.8 - \left[\frac{30(47.4) + 16(45.2) + 16(47.0)}{62}\right]$$

$$= 54.8 - 46.7$$

$$= 8.1$$

The squared standard error of $\hat{\Psi}_E$ is given by:

$$SE^2_{\hat{\Psi}_E} = MSW_E\left[\frac{1}{27} + \frac{1}{62}\right] = (7.2)^2\left[\frac{1}{27} + \frac{1}{62}\right] = 2.7553$$

so that:

$$t = \frac{8.1}{\sqrt{2.7553}} = 4.88$$

In this case, the hypothesis of no difference is not rejected since $t = 4.88 < 5.01$. Since the collective comparisons for fluency and dependent clauses were significant and since the three variables under consideration are strongly correlated within the groups, it makes sense to assume and conclude that the collective elaboration index difference does indeed represent a significant source of variance.

With respect to conventionality, it is clear that Group One differs from Groups Six, Seven, and Eight, individually, and therefore collectively. Since none of the maze differences involving Group One subjects is significant, it is concluded that Group One students continue to differ from Group Six, Seven, and Eight students on all of the language variables except freedom from mazes.

Group Two students are definitely unlike those of Groups Five, Six, Seven, and Eight with respect to conventionality. When they are compared collectively with the students in Groups Six, Seven, and Eight with respect to fluency, it is found that they are statistically different. For the fluency measure the appropriate contrast is given by:

$$\begin{aligned}\hat{\Psi}_F &= 54.3 - \left[\frac{30(46.2) + 16(45.3) + 16(44.0)}{62}\right] \\ &= 54.3 - 45.4 \\ &= 8.9\end{aligned}$$

The squared standard error of $\hat{\Psi}_F$ is given by:

$$SE^2_{\hat{\Psi}_F} = MSW_F\left[\frac{1}{51} + \frac{1}{62}\right] = (7.8)^2\left[\frac{1}{51} + \frac{1}{62}\right] = 2.1732$$

so that:

$$t = \frac{8.9}{\sqrt{2.1732}} = 6.01$$

Since $t = 6.01 > 5.01$, it is concluded that Group Two differs from the collective Group Six, Seven, and Eight with respect to fluency.

When the comparisons involving Group Three subjects are examined, it is seen that Group Three subjects differ from those in Group Six, Seven, and Eight with respect to conventionality but not with respect to any of the remaining language variables. Finally, it is seen that Group Four and Five subjects differ from Group Seven subjects with respect to conventionality.

If the statistical findings of the later ages are combined, it follows that Group One, Two, Three, Four, and Five subjects are statistically different from those in Group Six, Seven, and Eight with respect to conventionality. In addition, the subjects within the two sets of groups are not different from one another. Group One differs from Groups Six, Seven, and Eight with respect to fluency, dependent clauses, and elaboration index. Finally, Group Two differs from Groups Six, Seven, and Eight with respect to fluency.

Summary Comments on the Similarities and Differences Between the Language Groups at Grades Ten, Eleven, and Twelve

For the later years' data, only conventionality produces any pairwise differences that are statistically significant. However, when data for fluency, dependent clauses, and elaboration index are combined by groups, then a few other differences are detected.

With respect to conventionality, Group One students differ from Group Six, Seven, and Eight students. When the data for these groups are combined, it is also found that fluency, dependent clauses, and elaboration index are discriminators of some importance. With respect to conventionality, the mean difference exceeds 13.1 standardized units. For fluency, dependent clauses, and elaboration index, the corresponding mean differences are 10.8, 8.8, and 8.1 standardized units.

Group Two students differ from Group Five, Six, Seven, and Eight students with respect to conventionality. When Group Five is excluded from the collective group, it is found that Group Two students differ from the remainder with respect to fluency.

Finally, Group Three, Four, and Five students differ from Group Six, Seven, and Eight students with respect to conventionality.

Statistical Analysis of the Language Profiles Employing Linear Discriminant Functions

In some respects the multivariate analysis of the previous sections demanded close attention to details and an assimilation of many facts and findings to help decipher the information contained in the basic statistics and the differences between the means of the eight language groups for the five language variables. Some of this complexity can be reduced by examining the multivariate data by means of linear discriminant functions(18). In adopting this mode of analysis, detail is lost, but similar to the canonical correlation analysis, insight into the total overall language usage is achieved.

The basic principles involved in discriminant analysis are quite easy to understand and with the use of high speed computers their determination is easy to execute. For this analysis, the basic data consists of the 211 vectors of observations on the five language variables of each student in the study. For each vector of observations, the following linear compound, called a linear discriminant function, is created:

$$D = a_1X_1 + a_2X_2 + \dots + a_pX_p$$

Similar to principle components and canonical variates, the a_1, a_2, ..., a_p defining the compound are originally unspecified. Thus, before the D value can be computed for each individual subject, it is necessary to determine the values of the individual a_p.

For this determination, consider the G language groups and their average D values, $\overline{D}_1$, $\overline{D}_2$, ..., $\overline{D}_G$. From these values, the mean square between groups, MSB, can be computed by the following formula:

$$MSB = \frac{\sum_{g=1}^{G} n_G(\overline{D}_g - \overline{D})^2}{G - 1}$$

As would be expected, the value of MSB will depend upon the particular values assigned to the individual a_p. One strategy to employ in selecting the a_p is to choose the set of numbers that maximizes the MSB. The numbers that accomplish this maximization are intimately associated with the λ^{max} of the Roy Criteria used in the multivariate analysis of variance.

As is recalled, the λ^{max} of the Roy Criteria corresponds to the maximum root of the equation:

$$|B - \lambda W| = 0$$

This equation is called the characteristic equation of the system under analysis. One property of the characteristic equation is that, under general conditions, it has p roots called characteristic or eigen values. The Roy Criteria is based upon the maximum of these roots. In any case, the p roots can be ordered according to numerical value or size starting with λ^{max} and ending with λ^{min}. With each eigen value, there is an associated vector of constants. These constants define the p linear discriminant functions of the system. Thus, associated with a multivariate investigation involving p variables, are p linear discriminant functions. These functions can be tested for statistical significance. Those that are retained as significant can be used as hypothetical constructs to summarize the variables of the entire system in a manner similar to that of principle components or canonical variate analysis.

One other advantage to this procedure is that the discriminant functions that are retained are orthogonal. This means that if the underlying variables are multivariate normal, then the information contained in each discriminant function is independent of the information contained in the remaining functions. Operationally this means that each discriminant function can be treated as a hypothetical construct containing information independent of the information contained in the remaining hypothetical constructs.

The decision as to which functions to maintain is based upon a statistical test called Bartlett's Chi-square Test for Significance of Discriminant Functions(19). Those discriminant variates that are statistically significant can then be used as hypothetical constructs for a multivariate analysis of variance.

In this study, $p = 5$. To control the total probability of a Type I error to less than .05, each discriminant function is tested for significance with $\alpha = .01$. Those discriminant functions that are significant at this particular probability level are used in the remainder of this narrative to achieve a reduction of data.

For the first, second, and third grade data, the values of the individual a_p for the two statistically significant linear discriminant functions are as shown in Table 24. The first discriminant function or hypothetical construct is defined by:

$$D_1 = 12.1T_F - 4.5T_M + 10.0T_D + 4.8T_C + 9.8T_E$$

with:

$$\begin{aligned} Var(D_1) = {} & (12.1)^2Var(T_F) + (-4.5)^2Var(T_M) + \ldots + \\ & (9.8)^2Var(T_E) + 2(12.1)(-4.1)Cov(T_F, T_M) + \ldots + \\ & 2(4.8)(9.8)Cov(T_C, T_E) \\ = {} & 65105 \end{aligned}$$

The average value of D_1 is found by substituting $\overline{T}_F = \overline{T}_M = \overline{T}_D = \overline{T}_C = \overline{T}_E = 50$ into the equation for D_1. Thus, for these values of T_p, $D_1 = 1610$. By substituting the observed T scores for each of the variables for a particular subject into the equation for D, one can determine the D value for each student in the study. In addition, by substituting group average T scores into the same equation, the average D score for the group can be obtained with considerable ease.

Examination of the values of the constants that

Table 24. Linear Discriminant Functions for the First, Second, and Third Grade Data.

Variable	Discriminant Function One Raw Coefficient	Discriminant Function Two Raw Coefficient
Fluency	12.1	.6
Mazes	-4.5	10.8
Dependent Clauses	10.0	.8
Conventionality	4.8	10.9
Elaboration Index	9.8	-5.0
Average	1610	900
Variance	65105	13475
Standard Deviation	255.2	116.1

define D_1 show that it is primarily defined by fluency, dependent clauses, and elaboration index measures. Freedom from mazes and conventionality contribute less to the value of this hypothetical variable. Their lack of effect upon D_1 is demonstrated in Table 25 which shows the standardized D_1 values for various standardized T values of the five basic language variables. As can be seen, variation in mazes and conventionality leave the standardized scores quite unaffected even when they differ by two complete standard deviation units from the remaining three variables. Thus, a student with high fluency, high dependent clauses, and high elaboration index scores high on this hypothetical variable regardless of how high or low the scores on mazes and conventionality happen to be. On the other hand, a person who scores low on fluency, dependent clauses, and elaboration index, also scores low on this variable regardless of the score on mazes and conventionality. Since freedom from mazes and conventionality are minimal factors and since their impact is so small in the determination of D_1, it appears that D_1 is really a measure of language complexity and elaboration. Students who speak with elaborate and complex forms use language in a powerful way. Students at the opposite pole use language ineffectively. Thus, D_1 will be termed language power for the remainder of this narrative.

For the second discriminant function, which is orthogonal to the first, it is seen that fluency, dependent clauses, and elaboration index are almost entirely absent, so that the D_2 variable is defined by freedom from mazes and

Table 25. Interpretation of the First Linear Discriminant Score.

Variable	Case I	Case II	Case III	Case IV	Case V	Case VI	Case VII
Fluency	60	60	60	50	40	40	40
Mazes	60	50	40	50	60	50	40
Dependent Clauses	60	60	60	50	40	40	40
Conventionality	60	50	40	50	60	50	40
Elaboration Index	60	60	60	50	40	40	40
Discriminant Score	1932	1929	1926	1610	1294	1291	1288
Z Value	1.26	1.25	1.24	.00	-1.24	-1.25	-1.26
Standardized Score	62.6	62.5	62.4	50	37.6	37.5	37.4

conventionality. This second hypothetical construct is defined as:

$$D_2 = .6T_F + 10.8T_M + .8T_D + 10.9T_C - 5.0T_E$$

with:

$$\text{Var}(D_2) = 13475$$

Students who score high on freedom from mazes and conventionality score high on this variable independent of their scores on the remaining three variables. In like manner, students who score low on freedom from mazes and conventionality will also score low on D_2 and, what is more important, their score will be unaffected by their scores on the remaining three language variables. Students who are free from mazes and highly conventional in speech, speak with confidence. They are considerably unlike students who do not use standard speech, who hesitate, repeat themselves, and speak in loops because of lack of confidence. Therefore, this second measure will be termed language confidence.

In Table 26 are shown the standardized mean scores for the eight language groups on the two language variables defined by the two significant discriminant functions. In Figures 28 and 29 are shown the observed distributions of the standardized discriminant scores. If these figures are compared to Figures 9 through 14, the similarity of the graph for language power with fluency, dependent clauses, and elaboration index is striking. The same statements apply to the similarity of language confidence to freedom from mazes and conventionality. Thus, while five sets of figures were used to describe the basic data, the same information is essentially

Table 26. Basic Data for the Discriminant Scores for the Eight Language Groups at Grades One, Two, and Three.

Statistic	Group	D_1: Power	D_2: Confidence
Average	1	65.1	57.3
	2	52.9	56.7
	3	39.7	56.8
	4	52.4	49.3
	5	62.1	41.6
	6	43.5	43.0
	7	46.2	33.4
	8	33.6	42.5
Standard Deviation	1	5.6	4.3
	2	3.3	4.8
	3	4.4	4.1
	4	4.4	5.3
	5	5.9	5.4
	6	3.7	6.5
	7	6.2	10.6
	8	7.0	12.9
F-ratio		105.82*	41.21*

*Significant at $\alpha = .05$

Figure 28. Histograms of the Language Power Measures for the Eight Language Groups at Grades One, Two, and Three.

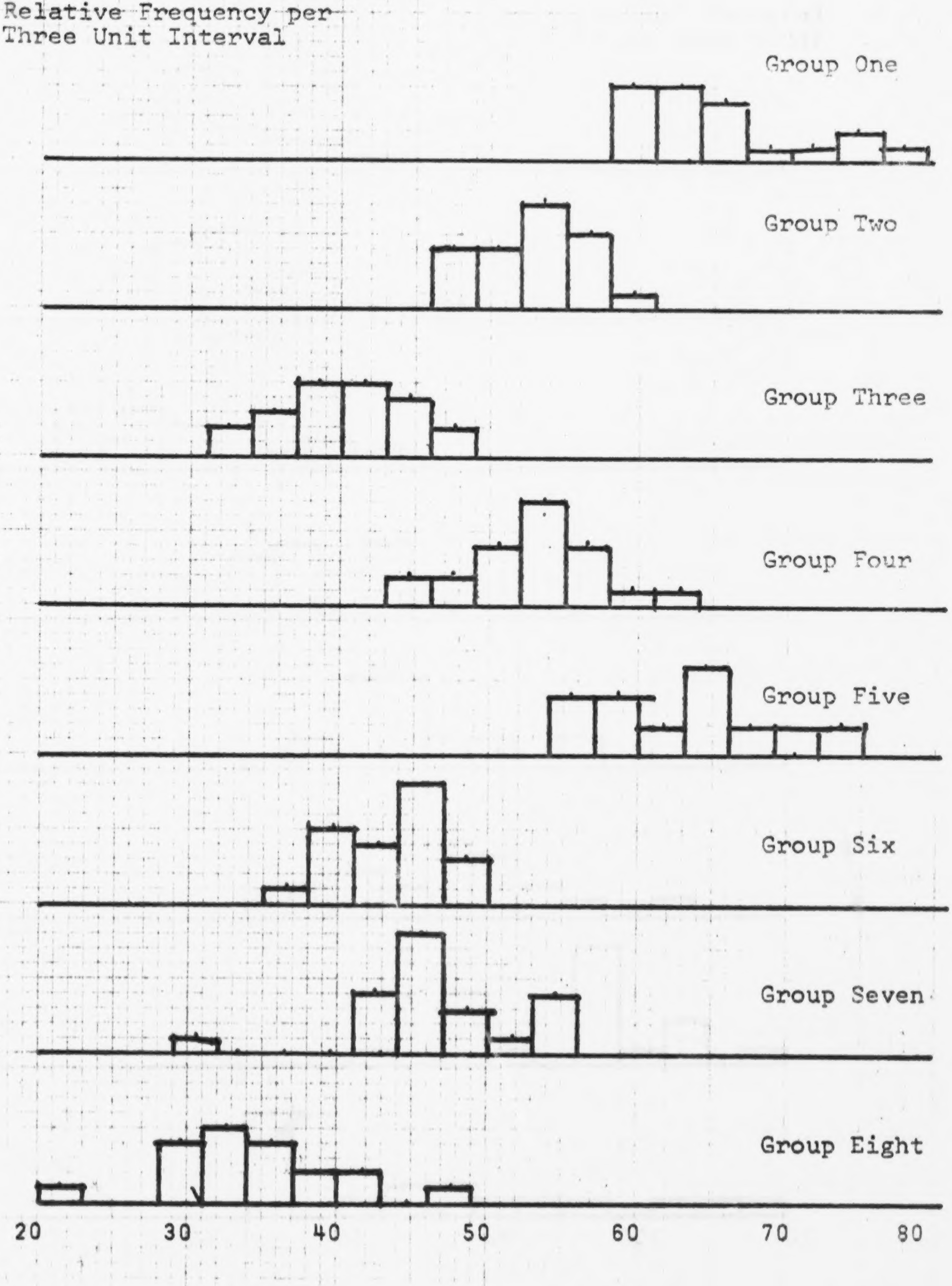
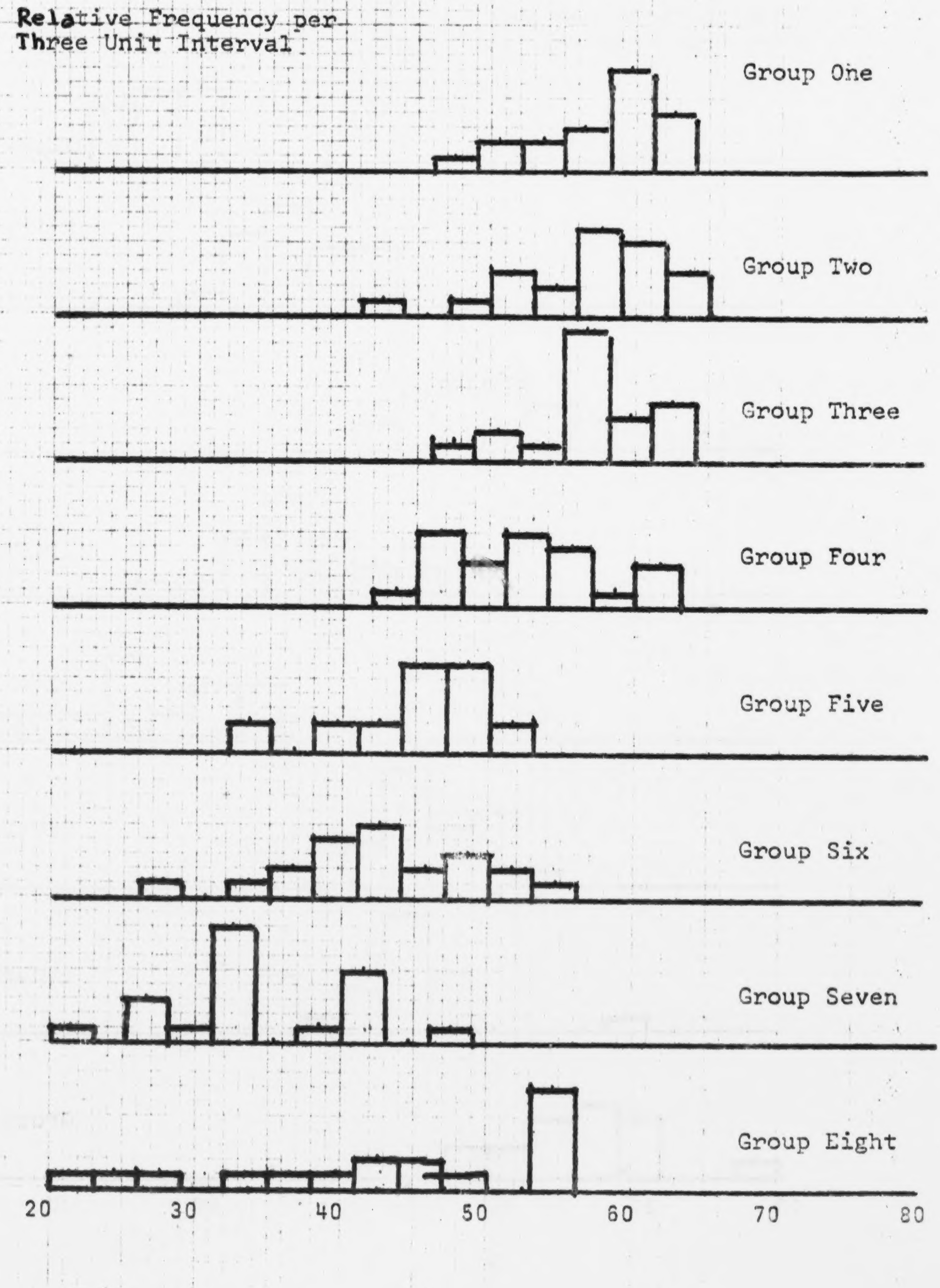


Figure 29. Histograms of the Language Confidence Measures for the Eight Language Groups at Grades One, Two, and Three.



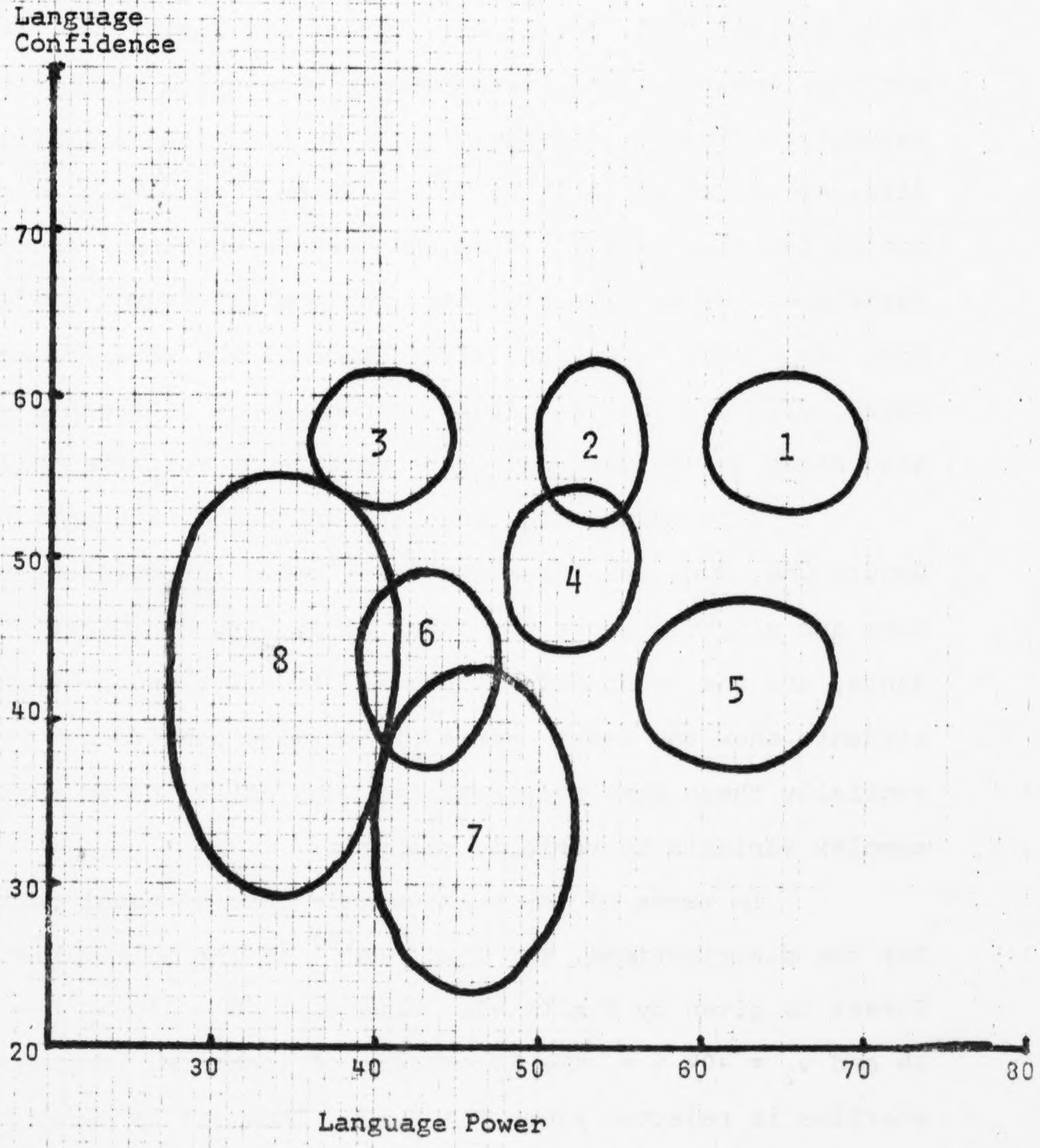
available in the two sets of graphs of Figures 28 and 29.

Figure 30 presents a graphic presentation of the bivariate distributions for the eight language groups. The individual ellipses represent the 68 percent central portion of the bivariate distributions for the eight language groups for the two discriminant functions. On the basis of language power, the hierarchy of the eight language groups is One, Five, Two and Four, Seven, Six, Three, and Eight. As noted earlier, Group One and Five students were quite similar with respect to fluency, dependent clauses, and elaboration index. Also, as was shown earlier, Group Two and Four students did not differ significantly from one another on these particular variables. Group Five students, who are all Negro, differ from Group Six, Seven, and Eight students who were primarily Negro, with respect to language power. This difference was also noted in the more detailed variable by variable analysis.

With respect to language confidence, students in Groups One, Two, and Three are identical. Intermediate to them and slightly above the Group Five, Six, and Eight students, are the Group Four students. Finally, Group Seven students show the least degree of language confidence. Essentially these same relationships were noted in the more complex variable by variable analysis.

In terms of the multivariate test of equal centers for the distributions, the Box form(3) of the multivariate F-test is given by $F = 64.41$. With $\nu_1 = (G - 1)p = (8 - 1)(2) = 14$ and $\nu_2 = 404 \sim \infty$, the hypothesis of identical language profiles is rejected since $F > 2.37$. Thus, it is clear that

Figure 30. The 68 percent Central Portions of the Bivariate Distributions for the Eight Language Groups Defined by Language Power and Language Confidence at Grades One, Two, and Three.



the language profiles on the two variables, language power and language confidence, are statistically different. As can be seen by examination of the univariate F-values shown in Table 26, both hypothetical variables are sources of statistical significance since $F = 105.82 > 2.08$ and $F = 41.21 > 2.08$.

As suggested by the mean values reported in Table 26, Groups One, Two, and Three are identical with respect to language power but do not differ from one another with respect to language confidence. Groups Six, Seven, and Eight are as noted earlier, the poorest users of language both with respect to confidence and power. Group Four students represent the median group of the study. Group Four students tend to use a slightly more complex form of language than the Group Five students, while the Group Five students tend to use language with slightly more confidence than the Group Four students.

For the Scheffé type analysis on the pairwise means, the critical values for the pairwise differences for the two linear discriminant variables are given by $C_{D_1} = 6.5$ and $C_{D_2} = 9.0$. Thus, any difference in means for D_1 exceeding 6.5 is significant while for D_2, any difference exceeding 9.0 is significant. The pairwise differences for the two language variables are shown in Table 27. As can be seen, Group One students differ from the remaining groups of students except Group Five students with respect to language power. With respect to language confidence, Group One does not differ from Groups Two, Three, and Four. Group Two students do not differ from Group Four students with respect to power nor do they differ from them with respect to confidence. This same identity of profiles was

Table 27. Discriminant Function Pairwise Comparisons Between the Profiles of the Eight Language Groups.

Group	1	2	3	4	5	6	7	8
Group One Versus the Seven Remaining Groups								
Power		12.4*	25.4*	12.7*	3.0	21.6*	18.9*	31.5*
Confidence		-.6	-.5	-8.0	-15.7*	-14.3*	-23.9*	-14.9*
Group Two Versus the Seven Remaining Groups								
Power			13.2*	.5	-9.3*	9.4*	6.7*	19.2*
Confidence			.1	-7.3	-15.1*	-13.7*	-23.3*	-14.3*
Group Three Versus the Seven Remaining Groups								
Power				-12.7*	-22.4*	-3.8	-6.5*	6.1
Confidence				-7.4	-15.2*	-13.8*	-23.4*	-14.4*
Group Four Versus the Seven Remaining Groups								
Power					-9.7*	8.9*	6.2	18.7*
Confidence					-7.7	-6.3	-15.9*	-6.9
Group Five Versus the Seven Remaining Groups								
Power						18.6*	15.9*	28.5*
Confidence						1.4	-8.2	.8
Group Six Versus the Seven Remaining Groups								
Power							-2.7	9.9*
Confidence							-9.6*	-.5
Group Seven Versus the Seven Remaining Groups								
Power								12.6*
Confidence								9.0

noted in the variable by variable analysis. Because the Group Three students tended to be nontalkers, they fail to differ from the Group Eight students with respect to power. This similarity was also noted for the variable by variable analysis. Group Four students, in addition to their similarity to Group Two students, are similar to Group Seven students, but since the mean difference of 6.2 is so close to the critical value of 6.5, it makes sense to declare the difference as significant. As noted earlier, Group Five students differ from all other groups except Group One students with respect to power. Finally, Group Six and Seven students are similar with respect to power. None of these similarities and differences are of any immediate surprise. Reasons for the differences were presented earlier and they still hold for this analysis. The histograms for D_3 are shown in Figure 30.

For the tenth, eleventh, and twelfth grade data, only one discriminant function is statistically significant. This discriminant function is given by:

$$D_3 = 2.9T_F + 1.0T_M + 3.0T_D + 15.4T_C - 1.6T_E$$

with:

$$\mathrm{Var}(D_3) = 14543$$

As this function indicates, the major discriminator between the eight language groups is conventionality. The coefficient of 15.4 for this variable tremendously outweighs the remaining coefficients in defining D_3 and so it makes sense to identify D_3 with language conventionality. The basic statistics for this one discriminant function is shown in Table 28.

As can be seen, the univariate F-ratio for this

Figure 30. Histograms of the Conventionality Measures for the Eight Language Groups at Grades Ten, Eleven, and Twelve.

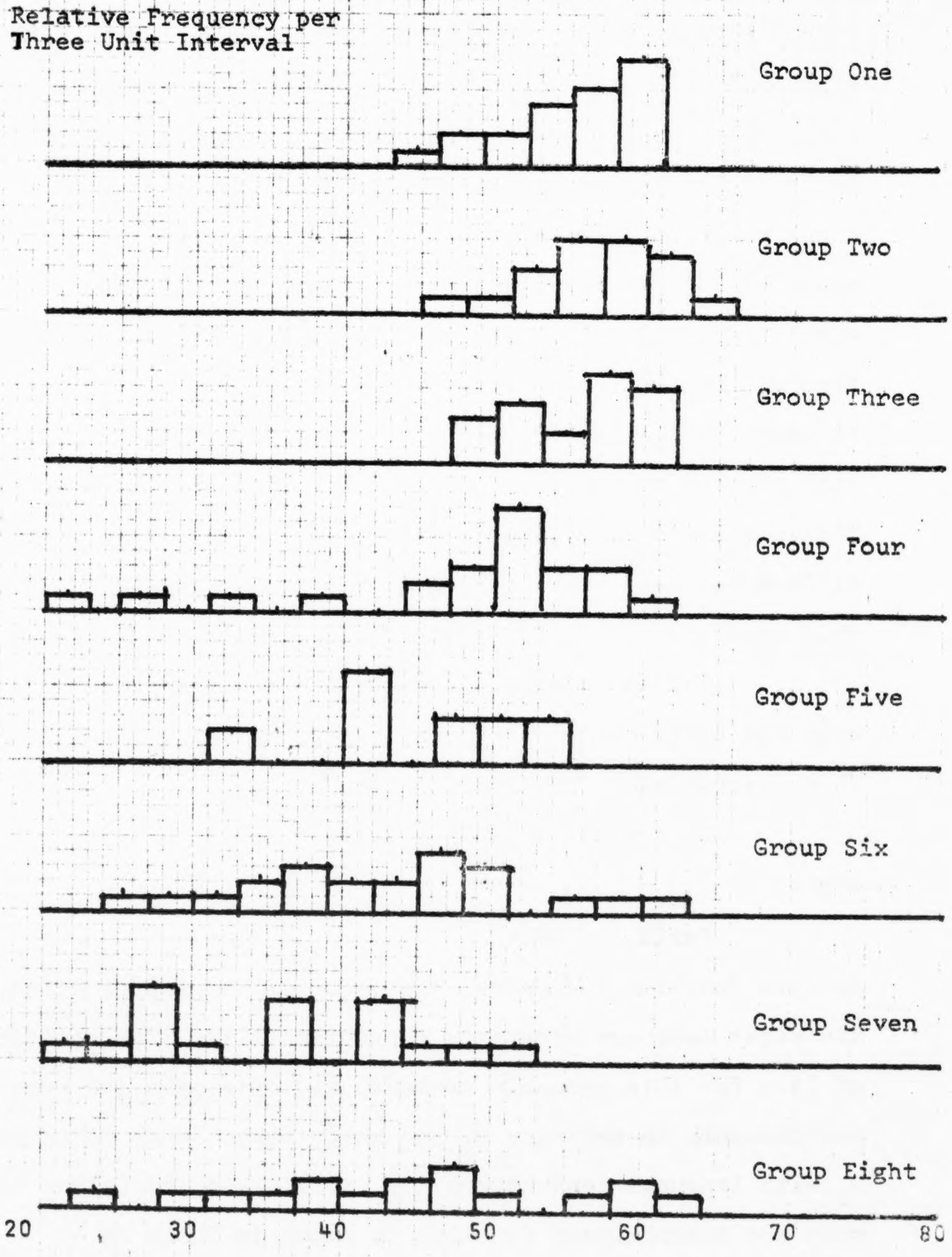


Table 28. Basic Statistics for the Discriminant Scores for the Eight Language Groups at Grades Ten, Eleven, and Twelve.

Statistic	Group	D_3: Conventionality
Average	1	55.7
	2	56.9
	3	55.3
	4	49.7
	5	46.4
	6	42.7
	7	35.0
	8	42.2
Standard Deviation	1	4.2
	2	4.2
	3	4.2
	4	8.7
	5	7.2
	6	9.1
	7	9.8
	8	11.8
F-ratio		27.47*

*Significant at $\alpha = .05$

variable is given by F = 27.47, which is considerably larger than 2.08. Thus, it is known that the mean differences between the groups are statistically significant. The pairwise differences are shown in Table 29. Any difference exceeding $C_{D_3} = 10.0$ represents a statistically significant difference. Exactly the same pairwise difference that were identified in Table 23 for the conventionality measures as statistically significant, are identified here as significant. There are no differences between the findings for the two sets of data.

Summary Findings Based on the Linear Discriminant Function Analysis

Linear discriminant functions were used in this report to simplify the analysis of the five variables across the eight language groups. The conclusions made on the basis of this analysis are identical to those reported earlier. The only difference is that they are presented in a more succinct manner.

Two significant linear discriminant functions were associated with the data generated at grades one, two, and three. They were labeled D_1: Language Power, and D_2: Language Confidence and were defined by the two orthogonal linear components:

$$D_1 = 12.1T_F - 4.5T_M + 10.0T_D + 4.8T_C + 9.8T_E$$

$$D_2 = .6T_F + 10.8T_M + .8T_D + 10.9T_C - 5.0T_E$$

With respect to D_1: Language Power, the hierarchy of students at the early ages was given by Group One, Group Five, Groups Two and Four, Group Seven, Group Six, Group Three, and

Table 29. Discriminant Function Pairwise Comparisons Between the Eight Language Groups.

Group	1	2	3	4	5	6	7	8
1		1.2	-.3	-6.0	-9.3	-13.0*	-20.7*	-13.5*
2			-1.5	-7.2	-10.5*	-14.2*	-20.9*	-14.7*
3				-5.6	-8.9	-12.6*	-20.3*	-13.2*
4					-3.3	-7.0	-14.7*	-7.5
5						-3.7	-11.4*	-4.2
6							-7.7	-.5
7								7.2

Group Eight. With respect to D_2: Language Confidence, the hierarchy is headed by Groups One, Two, and Three. Intermediate to them and slightly above Groups Five, Six, and Eight are the students of Group Four. Finally, Group Seven students close off the bottom of the listing.

For the tenth, eleventh, and twelfth grade data, only one discriminant function was statistically significant. This function was defined by:

$$D_3 = 2.9T_F + 1.0T_M + 3.0T_D + 15.4T_C - 1.6T_E$$

and is clearly a measure of conventionality suggesting that conventionality is the only variable that differentiates the groups from one another following elementary and secondary schooling. In terms of this variable, it is seen that Groups One, Two, and Three are equal. Next in the hierarchy is Group Four followed by Group Five, Groups Six and Eight, and finally, Group Seven.

Interpretations and Recommendations

The immediate questions that naturally arise following this analysis are what has caused the initial differences between the groups to be reduced or eliminated on all variables except conventionality and why have certain groups of students clustered together to form language styles and profiles that are not statistically different from one another.

As a possible explanation of these findings, it would appear that an upper limit exists to the amount of elaboration a speaker can impart when communicating in a verbal fashion. Students in Groups One, Two, Four, and Five were quite elaborate in their speech at an early age. For them to increase their relative standing with respect to elaboration at older ages is not to be expected. They started near the upper limits and could not be expected to improve appreciably over time. However, students who avoided elaborate speech at an earlier age have a long way to go before the upper limit is attained. Thus, as they mature and gain experience with language, the opprotunities to become more complex in speech and prone to detailing increases for them. Thus, one would expect students to come together over time with respect to this language characteristic. As is recalled, this convergence did occur. Group One students went from a mean standardized score of 61.9 to a mean standardized score of 54.8, while Group Eight students went from a mean standardized score of 35.8 to a mean standardized score of 47.0.

The same sort of convergence of language style would be expected for fluency and the use of dependent clauses because of their high intercorrelations. A child who makes effective use of dependent clauses at an earlier age will most likely continue to employ them as a means of expression with continuing age. On the other hand, a child who has avoided their use at an earlier age is not going to continue in this manner, but instead will be compelled to adopt them for speech and use them with greater frequency to express his thoughts and needs. Thus, formal classroom training in the structure and use of dependent clauses, multiple examples from written materials, continued exposure from teachers, classmates, friends, family, movies, records, and television are bound to have some transfer learning effects so that the use of these language components must increase for these students.

Finally, it is reasonable to expect mazes to decrease in frequency as facility with language increases and as vocabulary and skill in speaking improves and expands. This convergence was also noted for this characteristic. At grades one, two, and three, the range in average maze statistics extended from 36.3 to 57.3. At grades ten, eleven, and twelve, the range in average T-scores extended from 43.2 to 52.1.

Concerning the use of conventional language, it might be supposed that the same arguments could be used concerning the acquisition of standard, acceptable, conventional English. As a child is exposed to conventional speech at school, on television, and with friends, its use should automatically improve. However, the data of this investigation does not support that observation or those assumptions. Children who begin as the least conventional speakers remain the least conventional speakers. While their general ability to speak conventionally improves, the initial relative differences that exist between students remains. It appears that school, television, and other related language determiners are not overly effective in removing the initial state with which a student begins his spoken language history. A developing child who hears unconventional English at an early age continues to use the inappropriate models or habits throughout later life. It is quite conceivable that a lasting imprint is placed upon speech at an early age that is virtually impossible to erase. The children who hear conventional English from the start of speech learning are at a powerful language advantage in that as they grow linguistically, they are able to recognize unconventional forms of expression but not adopt them. The child whose language training and early experiences are not so conventional is at clear language development disadvantage. The dialect is implanted early and as long as it is reinforced, especially by parents and friends, its implant becomes stronger. This may be one of the major differences between conventional and unconventional language users. The conventional speaker learns the rules of acceptable speech early and does not adopt inappropriate models that appear as age and experience increase. On the other hand, the unconventional speaker has learned different rules and as a result, they continue as unconventional modes of expression at later ages. The general lack of change in conventionality over time for the eight groups of this study is demonstrated by the data of Table 30. As can be seen, mean conventionality scores at grades one, two, and three are almost identical to the conventionality scores at grades ten, eleven, and twelve. Clearly, these children have virtually maintained the rank order in conventionality of speech observed at the early time period of this study. In this sense, the hypothesis of no change in language has been supported.

What does the total collection of findings have to say about the basic null hypothesis of the study which essentially stated that language style does not change over

Table 30. Mean Conventionality Scores for the Eight Groups of Language Styles.

	Conventionality		
Group	Grades 1, 2, 3	Grades 10, 11, 12	Difference
1	55.8	56.1	.3
2	56.3	56.6	.3
3	55.4	54.1	-1.3
4	50.3	49.4	-.9
5	46.2	47.3	1.1
6	42.8	43.0	.2
7	39.7	36.7	-3.0
8	37.7	41.9	4.2

time. For the most part, the findings support this hypothesis with respect to conventionality of speech but definitely do not support the hypothesis with respect to problems with mazes. It essentially supports the hypothesis of no change with respect to fluency, dependent clauses, and elaboration index _for those youngsters who start out as poor users of spoken language._ These children improve their ability to speak effectively, but they never reach the proficiency of students who start out as good speakers.

The implications that these findings have for curriculum planning are pronounced, especially when it comes to the teaching of reading. Students who do not normally speak conventional English can not be expected to read with ease the text books that are conventional in the use of written language. Students who do not use conventional English should not be expected to comprehend, easily, written passages employing conventional English

In considering the implications of these findings for education, we will use for headings, "Diagnosis," "Method," "Content," and "Evaluation."

Diagnosis. Crucial to "Diagnosis" will be the measure of standard usage. Because conventionality scores for Groups One, Two, and Three are high and inasmuch as standard usage never proves to be a problem for members of these groups, teachers need to determine their position (in order to free them from any kind of drills or other school lessons aimed at teaching students the conventionality they already have or will acquire, both through interaction outside of school and in school.) To be sure, even in these three groups there may be some individuals who deviate so greatly from standard usage that they alone may need help.

Within any language group at the primary school level the correlations among the five language variables are low, close to zero in numerical value. This suggests that the five language variables are clearly providing unique measures on five different language characteristics within any one group of students. All five of them, therefore, are good diagnostic elements. To be sure, by the time all of these pupils have reached grades ten, eleven, and twelve--regardless of what group they are in--they will have moved much closer together on dependent clauses and elaboration of syntax measures. At this earlier level, however, only those in the higher groups are likely to use elaboration frequently and easily. Furthermore, there is considerable progress even for them on such matters as coiling dependent clauses into even tighter constructions. For instance, consider these three sets of communication units:

Less Mature, using dependent clauses

When Nina had fed the baby, she hurried after her father.

Literature is written so that it can clarify the real world.

The dog was in such a wild fury that he bit his master.

More Mature, using more tightly-coiled constructions.

Having fed the baby, Nina hurried after her father.
(present-perfect participial phrase)

Literature is writen to clarify the real world.
(infinitive phrase)

In his wild fury the dog bit his master.
(prepositional phrase)

The use of clauses reveals degrees of proficiency of language, but even more proficient is the ability to coil a dependent clause into more tightly constructed elaborations of thought as in the more mature sentences above. In the elementary school everyone, including the verbally competent child in Group One, has some distance to go yet in elaborating his communication units.

Since we know the language profiles of these groups are statistically different at the early ages, we recommend that these groups be identified in the early grades, for oral language is basic to the other language arts in that a spoken language is necessary before there can be any writing, any printing, or any reading. Identifying these language groups should make possible a more effective approach to instruction with these children. For instance, the children in Group Seven, coming from low socio-economic families with low measures of intelligence and a feeling of not being at home in the school environment, need special help on improving self-image. Diagnosis should alert teachers to the problem of damaged self-images possessed by this group of pupils.

The persistence of problems of non-standard usage and the disappearance of mazes over a twelve year period have definite implications. We see the importance of diagnosing the non-standard English and the child's receptivity to adding standard language to the dialect he already has.

Interest in mazes, however, need not be a part of diagnosis; eventually the problems of communication seem to iron out the mazes equally well in all students. Language diagnosis is time-consuming at best and teachers will easily notice, without diagnosis, those pupils who have extreme problems with mazes. No doubt a few will continue to have many loops and confusion in their oral expression; these might, of course, receive special speech therapy. However, great masses of children are never going to need help on this matter, and mazes at the senior high school level are no longer a significant source of variance among the groups. From the statistical evidence of this longitudinal study we can easily see that problems with mazes dissolve as students mature and learn to avoid distressing their listeners.

Since the five language variables measured at each of the two time periods give rise to a three-factor theory of language usage, we need to note that the three factors appear to be conventionality, freedom from mazes, and language power. Language power, of course, represents a combination of fluency, dependent clauses, and elaboration of syntax. For research purposes all three factors are useful. However, for classroom instruction we can agree that the problem of mazes can be eliminated, that the major problem is one of conventionality, and that language power will very likely be increased by methods having nothing to do with analysis of language but rather its use and manipulation.

Method. When we come to the matter of method, our speculation must center primarily on the matter of conventionality or the use of standard English. Here we relate our findings to what is already known about language learning. First of all, any method used should be oral and it should be one of use, of manipulation, of language to express ideas, of ear training, rather than of grammatical analysis. Language is not a subject like physics or mathematics. It is something used every day for functional and immediate purposes. The pupils do not need to learn language analysis, but rather involvement in discourse of all kinds is required. Students always learn through feedback as to whether or not they are communicating to their audiences the information they genuinely wish to communicate. Secondly, no method will succeed without motivation. Unless a pupil genuinely desires to add standard usage to whatever dialect or non-standard usage he may already have, any effort to influence his language will be futile. For example, many of the children in the group studied here are Negro. Some of them speak in non-standard ways. Among other planned experiences, they need language interchanges with standard speaking children and older people of their own ethnic group. These experiences should lead to an awareness of the prestige dialect without a loss of respect for the home dialect. Inasmuch as more than sixty percent of the non-conventional language in grades ten, eleven, and twelve in this research is already predictable in grades one, two, and three, it would seem important to have the very best instruction for the small children and a sustained program, both in guidance concerning language attitudes and in language learning, throughout all the years intervening between the elementary school and the high school. These children need a considerable amount of oral practice and much grappling with language problems in real communication situations.

Content. Teaching about language will mean creating situations where the students discover for themselves the differences between conventionality and non-conventionality, between powerful, elaborate syntax and weak, repetitive kernel sentences. The successful teacher will not impart such information, information the child would take in as a set of words only. The child will attain understanding and intellectual growth by discovering the structures of language. The teacher's function is to create the situations in which this learning can occur. This requires a shift from teaching a closed body of content in language arts, a body of content to be covered, to an open inquiry into language through which students propose questions and will have some prospect of developing their own answers.

Inasmuch as the students of Group Two need to speak with more complexity, if they are to gain greater fluency and power, these children will heed to discuss more complex matters than they have in the past, but the content must be something in which they are extremely interested. We would recommend for their curriculum, also, a considerable admixture of content dealing with language itself. That is, they might throughout the upper elementary school grades and the junior high schools study what is frequently called the miracle of language. There would be many lessons, delightful and interesting, dealing with the interest and fun of words and language. These children would be exposed to an interest in communication and how it is accomplished.

The children in Group Five, on the other hand, are already powerful in their use of syntax. The content of their language lessons would, of course, need to continue to be of deep interest to them and should be challenging to them in the sense that they would require complexity in the thoughts and feelings to be expressed. However, the particular problem is to help these children acquire the standard dialect in addition to their own dialect and to do so without damaging in any way their self-images.

We note that for Group Three, mostly girls and shy at the beginning, the greatest changes and improvements in language development take place. These children had high intelligence quotients and their initial language was conventional and free from mazes. The typical school instruction seems to work well for this type of child, and a model similar to what has been known in the past could be offered this group. Group Four also profits from typical educational instruction as we know it today; they have increased their ability to use standard speech, and the problem with mazes has disappeared over the years. These are typical children and they have responded to a typical education, a system largely designed to help them.

The real problems occur with the more gifted at the top of the language ladder and the less gifted at the bottom. For instance, the subjects in Group Six have not markedly enriched their language through schooling, maturation, or experience. And in Group Seven the subjects appear to have, over time, lost ground (relatively) with respect to all language variables except mazes. Both these groups of students were limited in language power to begin with, and obviously if the schools are to serve them, instruction must find new ways of dealing with them, ways different from those that succeed with Groups Three andFour. The emphasis upon oral language already described in this section represents our proposal of the best ways to help not only these children, but also that very special group known as Group Five as well as the groups at the very top of the language ladder. However, Groups Six and Seven do need attention to acquiring a second dialect, a standard dialect, whereas Groups One and Two do not need this. Group Five needs everything--a continued support of their syntactical ability and a subtle assistance in acquiring the standard dialect as an additional means of expressing themselves.

Very clearly the univariate analysis of variance for the five language variables shows that special attention to mazes need not be part of the curriculum for any of these groups. Apparently, what needs to be done about mazes is to focus a student's attention upon communicating to other people, of understanding how other people receive what he says. The egocentric quality of small children shows in their failure to take into account the difference between their own understanding of what they're saying and that of their listener's. Clearly, drills in language and a study of grammar are not going to help this at all.

Children need to be placed in situations where they must make their thoughts and feelings--and therefore their language--clear to someone else. Even in nursery school, with children who hold up a cup and say nothing, the teacher deliberately says, "What do you want?" When the child says, "More," she will often say, "More of what? Please tell me." This is to elicit a functional use of language. Children as they grow older should discuss situations in which they have failed to get a message, to receive directions, or to comprehend something. What went wrong? What are the observances we must always keep in mind when we are communicating to another person? Such a language program would be filled with functional uses of language where students are genuinely communicating to other people ideas and feelings they sincerely wish to communicate. Only in this way will they come to subordinate structurally those elements of thought or feeling which are subordinate semantically. Their transitions and all coherences should become increasingly logical in such situations. Teachers will need to watch to see whether they can sustain a line of thought and maintain a consistent point of view when they wish to or change a point of view when they wish to do that. At the senior high school level all these students will need to learn how to shift styles of language to suit different sorts of listener-decoders.

In the elementary school, children who find metaphors to capture the essentials of an intellectual or emotional meaning should be praised. Beginning somewhere in the junior high school and vigorously continuing in the senior high school should be a great deal of attention to the power of analogy, the use of metaphor. Back and forth to literature to composition to speech to listening, teachers should exalt the place of metaphor in language.

Grammar, as such, should be severely limited; no doubt it should come in as a small part of lessons which might be called the "miracle of language," for during any study of language some restrained attention to the structure of sentences is justified. The major emphasis would be upon subjects and predicates as the bases of structure for sentences, followed by the ways in which modification and coordination extend and elaborate these bases.

Evaluation. From this research we learn that in the elementary school each of our five language variables is a separate entity. In evaluating primary school children's language, then, one would wish to use measures of all five. Since very little evaluation of oral language has been accomplished in American education, or in world education, new devices such as the use of tapes and recorders will need to be introduced. Because this kind of evaluation is time-consuming, schools may need to select random samples for evaluation and employ teacher aides. Then teacher involvement will be in using the results of the evaluation, in choice of materials, equipment, and application of the findings.

At the tenth, eleventh, and twelfth grades evaluation may be done, however, mainly upon conventionality and fluency (length of communication unit). Mazes have dropped out of the picture, and fluency apparently does a good enough job of including under its rubric the problems of dependent clauses and elaboration of syntax. Evaluation of fluency in primary and upper elementary school, however, should take into account such matters as the ability of the child to coil dependent clauses into even tighter constructions such as appositives, nominative absolutes, gerund and participial clauses and phrases, and infinitives and prepositional phrases. Evaluation should also take into account the kinds of transitions used, as well as whether or not the child describes a coherent line of thought and maintains a consistent point of view.

More advanced evaluation should study the child in different situations with various audiences, audiences ranging from small children to adults, to see whether or not he is able to shift his styles flexibly in order to suit the different kinds of listeners and decoders of his speech. A very great amount of evaluation should examine tapes to see whether or not pupils use metaphor and analogy appropriately. Also, other thinking skills, such as ability to generalize, to give enough detail but not too much, to use both inductive and deductive thinking, to show sensitivity to the point of view, the thoughts and feelings of other people, to be able to detect false logic and propaganda, to be able to avoid narrow bias, to enter into a multiplicity and plurality of concepts of how different people think and how problems are solved, should also be a part of evaluation.

We have said nothing about vocabulary in this study. One reason for that is that vocabulary is extremely difficult to measure and time-consuming, and as yet the Loban research has processed very little on vocabulary beyond the first three grades. However, ultimately a great deal more needs to be known about vocabulary and should be included in the evaluation no matter if it is executed rather crudely. After all, what is evaluated tends to determine the curriculum, and anything that is left out of evaluation is likely to be left out of the curriculum. The power to evaluate is actually the power to control the curriculum.

In evaluation we need to remember that the top groups--Groups One, Two, Three, and Five--probably reach the highest levels of syntactical complexity somewhere in grades ten, eleven, and twelve and can go no further. There is research to show that about seven elements of information represent the most anyone can pack into a communication unit and still receive understanding from those who decode the unit. Evaluation should make a careful study of the amount of communication being coiled into a single unit by these exceptionally good speakers. Are they coiling a sufficient amount, but not too much?

All this preoccupation with oral language as the base for successful writing and reading--leading ultimately to appreciation of literature and to an awareness of language as a means of putting order into all of living--all this depends upon whether or not any transfer takes place. What is learned about speaking must have some valuable carry-over to writing and reading. We do not wish to be misunderstood; we realize writing and reading have conventions and domains of their own. That they cannot be acquired successfully without a base in oral language is the point. Here it is important to remember that learning equips a pupil with broad patterns of behavior rather than one-to-one relationships, that much of the waste in education results from workbook drills on details never consolidated into a comprehensive pattern. Let us look at some examples: the ability to write complete sentences or to read complicated sentences (using, for instance, appositives or other interpolated material between subject and predicate) will be initially learned from speaking such sentences or from listening to someone else speak them. Diagramming sentences or learning grammatical principles will never do for the larger reality; awareness of the patterns of sound linked to thought will transfer to writing and reading much more effectively than one-to-one items in drill. Sensitivity to standard usage is a deep pattern of attitudes and skills: concern that others will receive one's communication without distraction; a distaste for sloppiness and, therefore, distaste for whatever violates grammatical concord; relaxed self-respect that permits one to speak easily and naturally with attention focused on ideas. Teaching the skill of they were (instead of they was) is useless if the skill is not embedded in a total pattern of sensitivity to communication. A teacher who achieves a classroom where involvement, thought, and discussion prevail finds that time for drill can be reduced. Involvement, sensitivity to others, clarity of thought, and self-respect--these are what transfer, carrying with them the dependent components of appropriate usage.

The concept of transfer raises an unresolved issue. Does a learner benefit from conscious identification of goals in oral language? Many teachers assume so. They stress the idea of economy of endeavor; they believe that helping learners become aware of goals appropriate to their stage of oral language development can be accomplished without permanently disturbing the basic unselfconsciousness essential for ease and naturalness of speech. These teachers believe it possible to prevent learners from marking time in blind alleys of endeavor or forming, accidentally, habits delaying speech development.

Yet the issue is puzzling and far from resolved. Other teachers believe that "judiciously providing challenges will promote development," and they fear the school will promote a language self-consciousness upsetting to the naturalness of speech. They would place much greater emphasis upon building the child's self-image and offering him opportunities for success in speech situations focussed upon communicating material to someone he very much wants to interest. The child's delight in speech and his desire to use it effectively will outweigh, infinitely, any attempts to focus his mind upon how he talks. Explicit attention to improvement will accomplish nothing that would not develop naturally within genuine and varied communication experiences devised by the teacher.

Until we have more solid knowledge on this issue, we must select as reasonable a path as we can. Very likely the truth will prove to be some combination of the two positions. An effective teacher does both tasks, devising a wide variety of situations for natural informal talk and focussing attention on how improvement is possible. Many situations will emerge spontaneously from the interests and life of the classroom. Most will rise from drama or informal talk in small groups; others will take the form of round table or classroom discussion, still quite informal; and only a very small number will involve individual presentations before the class.

Nevertheless, for economical learning, the pupil must become aware of important rhetorical goals: the strategies of emphasis, the skills of exemplifying and generalizing, the importance of unity and relevancy. In the elementary years of schooling, pupils should merely be unconsciously aware of such goals, but in the secondary years the goals could become increasingly explicit. Selecting and learning the behaviors leading to these goals can be made more economical through teacher guidance, through models, and through motivated experience. The process is one of establishing goals the child understands and accepts. Teachers will need, of course, to have knowledge about language maturation and child development in order to avoid introducing goals too early or too late. The teacher's assistance refines and sharpens the learner's own observations and strategies for improvement.

Some ways oral language may be interwoven with other elements of the language program are suggested in the practices which follow. Though in their present form these suggestions do not carry all the implications, all the richness they would have in actual classrooms, they do illustrate the practices of teachers who believe in the linkage between spoken language and the other arts of communication.

With Writing

...Encourage pupils to write dramatic skits, act out what they have written, and revise this written form to as much perfection as possible.

...Show pupils how to read "jaberwocky" meaningfully; then help them transfer the skill to reading their own compositions.

...Expand skeletal sentences on the board, e.g., the coyote ate, using modification, compounding, cumulative clustering. The pupils read aloud, using vocal signalling to show meaning, then transfer the experience to their own writing.

...Establish in pupils the habit of looking back over their writing, of hearing with an "inner ear" how it might sound to another reader. To develop this "inner ear" each pupil reads aloud his own writing. This device helps young writers become more aware of ambiguities, awkward expressions, monotony of word choice and sentence pattern. Teachers may foster this habit in different ways:

a. Allow class time, before compositions are handed in, for each pupil to reread his work in a quiet voice.

b. Provide an audience by placing students in small groups or pairs to read their compositions.

c. Encourage each pupil to read into a tape recorder, then listen to his own voice as he follows his written form. In judging the style and tone of their own work, many writers find the ear a more reliable guide than the eye.

d. Keep model sentences on the chalkboard: cumulative sentences, compound and complex sentences, sentences with appositives, infinitive clauses. Read these aloud and discuss them. Have pupils choose model sentences from their own writing and place them on the chalkboard.

e. Teach manipulation of sentences; some elementary teachers begin by writing, each on a separate placard, the words of interesting sentences; then the word placards are distributed to pupils who come to the front of the room, arranging themselves according to directions from the class. Various alternatives of arranging the syntactical elements are tried for each sentence; the sentences are read aloud with various intonations and emphases. This exercise is then followed by similar seat work with words written on smaller cards. Pupils are encouraged to apply this manipulation to their own writing, using speech to test the various ways of arranging their own sentences.

With Reading

...Let initial reading instruction be a matter of helping the child make a transition from oral to written language efficiently and successfully, stressing the inductive learning of regular phoneme-grapheme patterns. Do not completely avoid irregularly spelled words but de-emphasize them during this phase, the language experience approach to reading. Later, exceptional words can be introduced in a controlled and gradual manner. At the beginning of reading, however, problems of word recognition should be reduced to a minimum. When children learn words, the words should be used orally in phrases and sentences so pupils become alert to the ways the words sound in the larger intonational setting.

...When he begins to read, the child should clearly see reading and writing as the reproduction of spoken language. This implies that beginning reading will use the dictation of children's language, both in individual records and in group experience charts. Sylvia Ashton-Warner(20) and Robert Van Allen(21) have described these methods fully enough for us to adopt them in our schools.

...Children should do much oral reading with the idea that they are to make their voices express the ideas just as they are expressed when spoken naturally. For teachers, parents, and children to remain content with the sing-song, colorless chants so often miscalled "reading" is dangerous. To be sure, a child may have to read silently, even practice aloud, before reading the living sound of language.

...Because children usually read aloud with a lack of meaningful intonation, tell them, "Good, first you need to be certain you recognize all the words. But now put the words together and read them as they should sound when you are speaking naturally."

...Let pupils take turns reading drama aloud. Strive for naturalness of tone, an imparting tone enhancing the meaning of the prose. (First let pupils read the material silently) Be sure the listening situation is motivated for such reading.

...Read sentences with varying patterns of intonation. The pupils imitate the teacher's pitch, pause, and stress. Apply the exercise to some functional use.

...Let poor readers listen to tapes of easy reading books as they follow the printed form with their eyes. The tapes should present skillful and powerful readings by expert trained voices.

...Older pupils prepare an oral reading of children's stories. They go to the primary grades to read aloud such books as *The Camel Took a Walk* or *The Little White Rabbit Who Wanted Red Wings*.

With Usage

...Present usage drill only to pupils who need a certain skill, such as "It doesn't" for "It don't." Drills are either taped or read aloud by the teacher; pupils listen, after instruction, in order to classify sentences as standard or (a few) as non-standard.

...Waste no time on such divided usage as *It is I* or *It is me*; *Who are you looking for?* or *For whom are you looking?* At most, such items should be noted as examples of how language changes and is changing. Spend time gained on more significant items, such as *He don't*, *He brung it*.

Usage represents the established oral language habits of an individual. Internalized by the child as he hears and imitates the speech of home and neighborhood, it is not a deliberate plan rationalized on a conscious level. It is quite different from grammar. Most of us can say "I want him to be my friend" without knowing *grammatically* that *him* is the subject of an infinitive, that the subjects of infinitives, quite illogically, are in the accusative rather than the nominative case. We can transform "A catcher's mitt was given to *him*" into "*He* was given a catcher's mitt" without recourse to grammatical knowledge. It is usage, not grammar, that all of us depend upon in such sentences--and in millions of other sentences we utter. Just as we learn to develop our usage through the ear, so too, if standard speech is to be learned, the way will be oral, through the ear, not through drillbooks or any version of grammar.

Grammar, the fascinating and careful analysis of the structure of a language--its sound structure, word structure, phrase and sentence structure--is too complexly indirect to help much with usage. Whenever usage learning occurs, the learning should be based upon oral methods, either repetition after the teacher (but only by those in the class who need the help--the others should be excused) or repetition from tapes similar to those used in language laboratories. The pupil must hear and say the standard form; he must not fill in blanks, underline printed forms, or memorize principles.

In societies organized for stability through caste and class, language has always been a major means of maintaining the status quo. Even in a fluid society such as ours, where individual worth and aspiration are intended to count for more than fortunate or unfortunate birth, language still operates to preserve status distinctions and remains a major barrier to crossing social lines. On attitudes concerning language, teachers can learn much from sociology. As stated by Cohen(22), "We fear lower class speech and are inclined to give it no quarter. The more precarious our social status in the higher classes--that is, the closer we are to the line that divides the middle from the lower classes or the more recent our ascent from the lower strata--the more insistent we are on the purity of our linguistic credentials."

Realizing that human worth cannot be measured by the language or dialect a man uses, teachers will be more likely to help children acquire a standard English without making them ashamed of their own language. Such acquisition--not "improvement"--is easier in situations where drill and directed efforts are oral, where they are linked to language expressing ideas, attitudes, and values of genuine concern to the learners. To improve language ability a pupil must apply whatever is studied to situations in which he has something to say, a deep desire to say it, and someone to whom he genuinely wants to say it.

REFERENCES

(1) Center for Research and Evaluation, American Institutes for Research, P. O. Box 1113, Palo Alto, California, 94302.

(2) Loban, Walter. The Language of Elementary School Children, Stock No. P56-70, The National Council of Teachers of English, 508 South Sixth Street, Champaign, Illinois, 1963.

(3) Loban, Walter. Problems in Oral English, Stock No. P56-105, The National Council of Teachers of English, 508 South Sixth Street, Champaign, Illinois, 1966.

(4) Loban, Walter. Language Ability: Grades Seven, Eight, and Nine, OE-30018, Cooperative Research Monograph No. 18 Catalog No. FS 5.230:30018, Superintendent of Documents, U. S. Government Printing Office, Washington, D. C., 1966.

(5) Tucker, Ledyard. "Learning Theory and Multivariate Experiment: Illustration by Determination of Generalized Learning Curves." In Catell, Raymond B. (Ed.), Handbook of Multivariate Experimental Psychology, Rand McNally, Chicago, 1966.

(6) Morrison, Donald F. Multivariate Statistical Methods, McGraw-Hill, New York, 1967. pp. 221-256.

(7) Bartlett, M. S. "The statistical significance of canonical correlation," Biometrika, Vol. 32, 1941, pp. 29-38.

(8) Friedmann, H. P. and Rubin, J. Journal of the American Statistical Association, Dec. 1967, Vol. 62, No. 320.

(9) Bock, R. D. "Programming univariate and multivariate analysis of variance," Technometrics, Vol. 5, 1963, pp. 95-117.

(10) Wilks, S. S. "Certain generalizations in the analysis of variance," Biometrika, Vol 24, 1932, pp. 471-494.

(11) Box, G. E. P. "A general distribution theory for a class of likelihood criteria," Biometrika, Vol 36, 1946, pp. 317-346.

(12) Kaiser, H. F. "The varimax criterion for analytic rotation in factor analysis," Psychometrika, Vol 23, 1958, pp. 187-200.

(13) Tryon, R. C., and Barley, D. E. "The BCTRY computer system of cluster and factor analysis," Multivariate Behavior Research, Vol. 1, 1966, pp. 95-111.

(14) Hays, William L. Statistics for Psychologists, Holt, Rinehart, and Winston, New York, 1963.

(15) Scheffé, Henry. "A method for judging all contrasts in the analysis of variance," Biometrika, Vol. 40, 1953, pp. 87-104.

(16) Heck, D. L. "Charts of some upper percentage points of the distribution of the largest characteristic root," Annals of Math Stat, Vol. 31, 1960, pp. 625-642.

(17) Roy, S. N., and Bose, R. C. "Simultaneous confidence interval estimation," Annals of Math Stat, Vol. 24, 1953, pp. 513-536.

(18) Jones, Lyle. Analysis of variance in its multivariate developments. In Cattell, Raymond B., (ed.), Handbook of Multivariate Experimental Psychology, Rand McNally, 1966, Chicago.

(19) Bartlett, M. S. "The goodness of fit of a single hypothetical discrimination function in the case of several groups," Annals of Eugenics, 1951, Vol. 16, pp. 199-214.

(20) Ashton-Warner, Sylvia. Teacher, Simon and Schuster, New York, 1963.

(21) Lee, Doris and Van Allan, R. Learning to Read Through Experience, Appleton Century, New York, 1963.

(22) Cohen, Werner. "On the language of lower class children," School Review, Winter 1959, pp. 35-40.